

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2003, 12:38:09 ; Search time 80 Seconds  
(without alignments)  
816.462 Million cell updates/sec

Title: US-10-083-853B-1  
Perfect score: 1704  
Sequence: 1 MGDFMSKTPKAMATKAKID.....DIQLDHERCDICSSRPPR 317

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

SPREMBL 21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description
1	678.5	39.8	641	Q14288	Q14288 homo sapien
2	678	39.8	1275	Q00362	Q00362 homo sapien
3	676.5	39.7	1275	Q00363	Q00363 homo sapien
4	674.5	39.6	1275	Q095K0	Q095K0 homo sapien
5	674.5	39.6	1275	Q08TE30	Q08TE30 homo sapien
6	674	39.6	712	Q14754	Q14754 homo sapien
7	673.5	39.5	1275	Q00370	Q00370 homo sapien
8	673.5	39.5	1275	Q00372	Q00372 homo sapien
9	672.5	39.5	1275	Q15604	Q15604 homo sapien
10	672.5	39.5	1275	Q12881	Q12881 homo sapien
11	672.5	39.5	1275	Q00360	Q00360 homo sapien
12	672.5	39.5	1275	Q00366	Q00366 homo sapien
13	672.5	39.5	1275	Q00368	Q00368 homo sapien
14	672.5	39.5	1275	Q09UN80	Q09UN80 homo sapien
15	672.5	39.5	1275	Q15606	Q15606 homo sapien
16	670.5	39.3	1275	Q00375	Q00375 homo sapien

17	661.5	38.8	1275	4	Q00378	Q00378 homo sapien
18	637	37.4	364	4	Q9BWM5	Q9BWM5 homo sapien
19	637	37.4	370	4	Q9GZP0	Q9GZP0 homo sapien
20	588	34.5	290	11	Q9P1L8	Q9P1L8 mus musculu
21	588	34.5	370	11	Q9Z5I7	Q9Z5I7 mus musculu
22	585	34.3	370	11	Q9EQ11	Q9EQ11 mus musculu
23	463.5	27.2	646	11	Q63779	Q63779 rattus norv
24	463.5	27.2	1300	11	P97692	P97692 rattus norv
25	425.5	25.0	275	11	Q63306	Q63306 rattus norv
26	425	24.9	1275	6	Q6Z5S8	Q6Z5S8 canis fami1
27	413.5	24.3	1252	11	Q9QWY2	Q9QWY2 mus musculu
28	413.5	24.3	1281	11	Q88913	Q88913 mus musculu
29	413.5	24.3	1281	11	Q88914	Q88914 mus musculu
30	413.5	24.3	1281	11	Q9QWY3	Q9QWY3 mus musculu
31	413.5	24.3	1281	11	Q9QWY0	Q9QWY0 mus musculu
32	413.5	24.3	1281	11	Q88915	Q88915 mus musculu
33	413.5	24.3	1281	11	Q9QW12	Q9QW12 mus musculu
34	413.5	24.3	1281	11	Q54850	Q54850 mus musculu
35	413.5	24.3	1300	11	Q08906	Q08906 mus musculu
36	412.5	24.2	1281	11	Q60713	Q60713 mus musculu
37	412.5	24.2	1281	11	Q91Z89	Q91Z89 mus musculu
38	412.5	24.2	1281	11	Q91Z88	Q91Z88 mus musculu
39	412.5	24.2	1300	11	Q61785	Q61785 mus musculu
40	409.5	24.0	1219	11	Q9QYM3	Q9QYM3 mus musculu
41	389.5	22.9	367	11	Q63778	Q63778 rattus norv
42	374	21.9	392	6	Q28360	Q28360 didelphis m
43	359	21.1	136	4	Q9HAG0	Q9HAG0 homo sapien
44	344	20.2	92	4	Q14269	Q14269 homo sapien
45	341	20.0	246	4	Q96GQ3	Q96GQ3 homo sapien

## ALIGNMENTS

### RESULT 1

ID Q14288 PRELIMINARY; PRT; 641 AA.  
AC Q14288;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Hypothetical 75.3 kDa protein (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Woods-Samuels P.;  
RL Submitted (APR-1989) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=8923117; PubMed=2497061;  
RA Kazanian H.H. Jr., Antonarakis S.E.;  
RT "Characterization of a nondeletitious L1 insertion in an intron of the human factor VIII gene and further evidence of open reading frames in functional L1 elements."  
RT Genomics 4:290-296(1989).  
RL EMBL; M22334; AAA88038.1; -  
DR InterPro; IPR000477; RVTse.  
DR Pfam; PF00078; rvc; 1  
KW Hypothetical protein; RNA-directed DNA polymerase.  
FT NON\_TER  
FT 1  
SQ SEQUENCE 641 AA; 75313 MW; 0A88757E7E3F66BF CRC64;

Query Match 39.8%; Score 678.5; DB 4; Length 641;  
Best Local Similarity 71.8%; Pred. No. 6,9e+48;  
Matches 148; Conservative 4; Mismatches 27; Indels 27; Gaps 5;

QY 1 MGDFMSKTPKAMATKAKIDKMDLILKSFCTAKETTFIVNQLTWEKIFATYSPDKGL 60  
DB 356 VGRDFMSKTPKAMATKAKIDKMDLILKSFCTAKETTFIVNQPTWEKIFATYSSDKGL 415

```

QY 61 ISRIYNELKOIYKKKTKNPDKKWKDMNRHPSKEGIYAAKKGMKKYSSSLAIREMOIKTT 120
DB 416 ISRIYNELKOIYKKKTKNPDKKWKDMNRHPSKEGIYAAKKGMKKYSSSLAIREMOIKTT 475
QY 121 MRHYLTPVRMAIIKKSGNN--RDMDAGNHHSQOTITRTKQTPHYLTHRW---110QS 174
DB 476 MRHYLTPVRMAIIKKSGNNRCMRGCEIG-----TLHGMWCKLVOP 518
QY 175 HWTV--LSDISELMHKTDRIVNLL 197
DB 519 LMKSVWFRLDRL-ELIEIPFDPALPL 543

RESULT 2
000362 PRELIMINARY; PRT; 1275 AA.
ID 000362;
AC 000362;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative p150.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97285120; PubMed=9140393;
RA Saasaman D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,
RA Deberardinis R.J., Gabriel A., Swergold G.D., Kazazian H.H. Jr.;
RT "Many human LI elements are capable of retrotransposition.";
RL Nat. Genet. 16:37-43(1997).
DR EMBL; U93564; AAC51263.1; -
DR InterPro; IPR005135; Exo_endo_phos.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR Pfam; PF00078; rvt; 1.
KW RNA-directed DNA polymerase.
SQ SEQUENCE 1275 AA; 149201 MW; 23D51D6B4358F28 CRC64;

Query Match 39.8%; Score 678; DB 4; Length 1275;
Best Local Similarity 94.3%; Pred. No. 1.7e-47;
Matches 132; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 MGKDFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQLTWEMKIFATYSPDKGL 60
DB 990 VGKDFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQLTWEMKIFATYSPDKGL 1049
QY 61 ISRIYNELKOIYKKKTKNPDKKWKDMNRHPSKEGIYAAKKGMKKYSSSLAIREMOIKTT 120
DB 1050 ISRIYNELKOIYKKKTKNPDKKWKDMNRHPSKEGIYAAKKGMKKYSSSLAIREMOIKTT 1109
QY 121 MRHYLTPVRMAIIKKSGNNR 140
DB 1110 MRHYLTPVRMAIIKKSGNNR 1129

RESULT 3
000363 PRELIMINARY; PRT; 1275 AA.
ID 000363;
AC 000363;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative p150.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97285120; PubMed=9140393;

```

```

RA Saasaman D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,
RA Deberardinis R.J., Gabriel A., Swergold G.D., Kazazian H.H. Jr.;
RT "Many human LI elements are capable of retrotransposition.";
RL Nat. Genet. 16:37-43(1997).
DR EMBL; U93565; AAC51264.1; -
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR Pfam; PF00078; rvt; 1.
KW RNA-directed DNA polymerase.
SQ SEQUENCE 1275 AA; 149008 MW; A866976A3FD8F74 CRC64;

Query Match 39.7%; Score 676.5; DB 4; Length 1275;
Best Local Similarity 71.8%; Pred. No. 2.3e-47;
Matches 148; Conservative 4; Mismatches 27; Indels 27; Gaps 5;

QY 1 MGKDFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQLTWEMKIFATYSPDKGL 60
DB 990 VGKDFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQLTWEMKIFATYSPDKGL 1049
QY 61 ISRIYNELKOIYKKKTKNPDKKWKDMNRHPSKEGIYAAKKGMKKYSSSLAIREMOIKTT 120
DB 1050 ISRIYNELKOIYKKKTKNPDKKWKDMNRHPSKEGIYAAKKGMKKYSSSLAIREMOIKTT 1109
QY 121 MRHYLTPVRMAIIKKSGNN--RDMDAGNHHSQOTITRTKQTPHYLTHRW---110QS 174
DB 1110 MRHYLTPVRMAIIKKSGNNRCMRGCEIG-----TLHGMWCKLVOP 1152
QY 175 HWTV--LSDISELMHKTDRIVNLL 197
DB 1153 LMKSVWFRLDRL-ELIEIPFDPALPL 1177

RESULT 4
09Y5K0 PRELIMINARY; PRT; 1275 AA.
ID 09Y5K0;
AC 09Y5K0;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical 149.0 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Divoky V., Indra K., Mrug M., Brabec V., Huisman T.H.J., Prchal J.T.;
RT "A novel mechanism of B-Chalasemia. The insertion of LI
RT retrotransposable element into B globin IVSII.";
RL Blood 88:148-148(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Kimberland M.L., Divoky V., Prchal J., Schwahn U., Berger W.,
RA Kazazian H.H.;
RT "Full-length human LI insertions retain the capacity for high
RT frequency retrotransposition in cultured cells.";
RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF149422; AAD38785.1; -
DR InterPro; IPR005135; Exo_endo_phos.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR Pfam; PF00078; rvt; 1.
KW Hypothetical protein, RNA-directed DNA polymerase.
SQ SEQUENCE 1275 AA; 149015 MW; 7E4AB526015ED37C CRC64;

Query Match 39.6%; Score 674.5; DB 4; Length 1275;
Best Local Similarity 71.4%; Pred. No. 3.3e-47;
Matches 147; Conservative 4; Mismatches 28; Indels 27; Gaps 5;

QY 1 MGKDFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQLTWEMKIFATYSPDKGL 60
DB 990 VGKDFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQLTWEMKIFATYSPDKGL 1049

```

```

QY 61 ISRIYNELKOIYKKTNPDKKVVKDMNRHFSKEGIYAACKHMKYSSSLAIREMOIKTT 120
DB 1050 ISRIYNELKOIYKKTNPDKKVVKDMNRHFSKEDIYAACKHMKYSSSLAIREMOIKTT 1109
QY 121 MRHYLTPVRMAIIKKSGNN---RDMDEAGNHHSOQITRTKNOTPHYLTRM---ILQOS 174
DB 1110 MRHYLTPVRMAIIKKSGNNRCWRCGCEIG-----TLHCHWDCKLVOP 1152
QY 175 HWYTV---ISDISELMHKTDRIVNLL 197
DB 1153 LMKSVWRFLRDL-ELRIPPDPALPL 1177

RESULT 5
Q8TE30
ID 08TE30 PRELIMINARY; PRT; 1275 AA.
AC 08TE30;
DT 01-JUN-2002 (TEMBLrel. 21, Created)
DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Hypoetical 149.0 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21668188; PubMed=11810275;
RA Benjes S.M., Morris C.M.;
RT "A full-length and potentially active LINE element is integrated
RT polymorphically within the IGL locus in a genomically unstable region
RT of chromosome 32."
RL Hum. Genet. 109:628-637(2001).
DR EMBL; AF421375; AAL50637.1; -.
KW Hypoetical protein.
SQ SEQUENCE 1275 AA; 149010 MW; B327D9D50A581764 CRC64;

Query Match 39.6%; Score 674.5; DB 4; Length 1275;
Best Local Similarity 71.4%; Pred. No. 3.3e-47;
Matches 147; Conservative 4; Mismatches 28; Indels 27; Gaps 5;

QY 1 MGDFMSKTPKMATATKIDKMDLIKLSFCTAKETTRIVNROLTEWEXIFATYSPDKGL 60
DB 990 VGKDFMSKTPKMATATKIDKMDLIKLSFCTAKETTRIVNROPTWEKIFATYSSDKGL 1049
QY 61 ISRIYNELKOIYKKTNPDKKVVKDMNRHFSKEGIYAACKHMKYSSSLAIREMOIKTT 120
DB 1050 ISRIYNELKOIYKKTNPDKKVVKDMNRHFSKEDIYAACKHMKYSSSLAIREMOIKTT 1109
QY 121 MRHYLTPVRMAIIKKSGNN---RDMDEAGNHHSOQITRTKNOTPHYLTRM---ILQOS 174
DB 1110 MRHYLTPVRMAIIKKSGNNRCWRCGCEIG-----TLHCHWDCKLVOP 1152
QY 175 HWYTV---ISDISELMHKTDRIVNLL 197
DB 1153 LMKSVWRFLRDL-ELRIPPDPALPL 1177

RESULT 6
Q14754
ID 014754 PRELIMINARY; PRT; 712 AA.
AC 014754;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE OSFII.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC TISSUE=PLACENTA;
RX MEDLINE=90323398; PubMed=2165587;
RA Hohjoh H., Minakami R., Sakaki Y.;
RT "Selective cloning of the human Ll (line-1) sequence which transposed
RT in a relatively recent past."
RL Nucleic Acids Res. 18:4099-4104(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=92319645; PubMed=1320255;
RA Minakami R., Kurase K., Etoh K., Furuhata Y., Hattori M., Sakaki Y.;
RT "Identification of an internal cis-element essential for the human Ll
RT transcription and a nuclear factor(s) binding to the element."
RL Nucleic Acids Res. 20:3139-3145(1992).
DR EMBL; X52235; CAA36480.1; -.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF00078; RVC; 1.
KW RNA-directed DNA polymerase.
SQ SEQUENCE 712 AA; 83610 MW; 9123697871E4599F CRC64;

Query Match 39.6%; Score 674; DB 4; Length 712;
Best Local Similarity 93.6%; Pred. No. 1.8e-47;
Matches 131; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 MGDFMSKTPKMATATKIDKMDLIKLSFCTAKETTRIVNROLTEWEXIFATYSPDKGL 60
DB 427 MGDFMSKTPKMATATKIDKMDLIKLSFCTAKETTRIVNROPTWEKIFATYSSDKGL 486
QY 61 ISRIYNELKOIYKKTNPDKKVVKDMNRHFSKEGIYAACKHMKYSSSLAIREMOIKTT 120
DB 487 ISRIYNELKOIYKKTNPDKKVVKDMNRHFSKEDIYAACKHMKYSSSLAIREMOIKTT 546
QY 121 MRHYLTPVRMAIIKKSGNNR 140
DB 547 MRHYLTPVRMAIIKKSGNNR 566

RESULT 7
Q00370
ID 000370 PRELIMINARY; PRT; 1275 AA.
AC 000370;
DT 01-JUL-1997 (TEMBLrel. 04, Created)
DT 01-JUL-1997 (TEMBLrel. 04, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Putative p150.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97285120; PubMed=9140393;
RA Sasseman D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,
RA DeBerardinis R.J., Gabriel A., Swergold G.D., Kazanian H.H. Jr.;
RT "Many human Ll elements are capable of retrotransposition."
RL Nat. Genet. 16:37-43(1997).
DR EMBL; U93569; AAC51271.1; -.
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR Pfam; PF00078; RVC; 1.
KW RNA-directed DNA polymerase.
SQ SEQUENCE 1275 AA; 149011 MW; 588703688E7129FF CRC64;

Query Match 39.5%; Score 673.5; DB 4; Length 1275;
Best Local Similarity 71.4%; Pred. No. 4e-47;
Matches 147; Conservative 4; Mismatches 28; Indels 27; Gaps 5;

QY 1 MGDFMSKTPKMATATKIDKMDLIKLSFCTAKETTRIVNROLTEWEXIFATYSPDKGL 60
DB 990 VGKDFMSKTPKMATATKIDKMDLIKLSFCTAKETTRIVNROPTWEKIFATYSSDKGL 1049
QY 61 ISRIYNELKOIYKKTNPDKKVVKDMNRHFSKEGIYAACKHMKYSSSLAIREMOIKTT 120

```

```

Db 1050 ISRIYNELKQIYKKTNNPIKKAQKDMNRHFSKEDIYAAKGMKCCSSLAIREMOIKTT 1109
Qy 121 MRHYLTPVRMAIIKKSGNN---RDMDEAGNHSSQOTITRTKNGQTPHYLTHRW---110QS 174
Db 1110 MRHYLTPVRMAIIKKSGNNRCWRGCGEIG-----TLHGWMDCKLVOP 1152
Qy 175 HWTV---LSDISELMHKTDRIVNLL 197
Db 1153 LKSWVRFLRDL-ELFIPDPALPL 1177

```

## RESULT 8

```

000372 ID 000372 PRELIMINARY; PRT; 1275 AA.
AC 000372;

```

```

Dt 01-JUL-1997 (TREMBlrel. 04, Created)
Dt 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
Dt 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative p150.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=97285120; PubMed=9140393;

```

```

RA Saasman D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,
RA Debernardi R.J., Gabriel A., Swergold G.D., Kazanian H.H. Jr.;
RT "Many human L1 elements are capable of retrotransposition.";
RL Nat. Genet. 16:37-43(1997).
DR EMBL: U93570; AAC51273.1;
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR004477; RVTse.
DR Pfam; PF03372; Exo_endo_phos. 1.
DR Pfam; PF00078; rvt. 1.
KM RNA-directed DNA polymerase.
SQ SEQUENCE 1275 AA; 149111 MW; 4711B3BC22F674E CRC64;

```

```

Query Match 39.5%; Score 673.5; DB 4; Length 1275;
Best Local Similarity 71.4%; Pred. No. 4e-47;
Matches 146; Conservative 5; Mismatches 31; Indels 21; Gaps 4;

```

```

Qy 1 MGDFMSKTPKAMATRAKIDKMDILKLSFCTAKETITRVNROLTEWEXIFATYSPDKGL 60
Db 990 VGKDFMSKTPKAMATRAKIDKMDILKLSFCTAKETITRVNROLTEWEXIFATYSPDKGL 1049
Qy 61 ISRIYNELKQIYKKTNNPIKKAQKDMNRHFSKEDIYAAKGMKCCSSLAIREMOIKTT 120
Db 1050 ISRIYNELKQIYKKTNNPIKKAQKDMNRHFSKEDIYAAKGMKCCSSLAIREMOIKTT 1109
Qy 121 MRHYLTPVRMAIIKKSGNNRCWRGCGEIG-----TLHGWMDCKLVOP 1152
Db 1110 MRHYLTPVRMAIIKKSGNNRCWRGCGEIG-----TLHGWMDCKLVOP 1152
Qy 178 TV---LSDISELMHKTDRIVNLL 197
Db 1156 SWVRFLRDL-ELFIPDPALPL 1177

```

## RESULT 9

```

015604 ID 015604 PRELIMINARY; PRT; 1275 AA.
AC 015604;

```

```

Dt 01-NOV-1996 (TREMBlrel. 01, Created)
Dt 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
Dt 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE ORF2 contains a reverse transcriptase domain.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RA Dombroski B.A.;
RT "Isolation of an active human transposable element.";
RL Science 0:0-0(1991).
DR EMBL; M80340; AAA51622.1;
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR004477; RVTse.
DR Pfam; PF03372; Exo_endo_phos. 1.
DR Pfam; PF00078; rvt. 1.
KM RNA-directed DNA polymerase.
SQ SEQUENCE 1275 AA; 149044 MW; C090473BBB0A44C1 CRC64;

```

```

Query Match 39.5%; Score 672.5; DB 4; Length 1275;
Best Local Similarity 71.4%; Pred. No. 4.8e-47;
Matches 147; Conservative 4; Mismatches 28; Indels 27; Gaps 5;

```

```

Qy 1 MGDFMSKTPKAMATRAKIDKMDILKLSFCTAKETITRVNROLTEWEXIFATYSPDKGL 60
Db 990 VGKDFMSKTPKAMATRAKIDKMDILKLSFCTAKETITRVNROLTEWEXIFATYSPDKGL 1049
Qy 61 ISRIYNELKQIYKKTNNPIKKAQKDMNRHFSKEDIYAAKGMKCCSSLAIREMOIKTT 120
Db 1050 ISRIYNELKQIYKKTNNPIKKAQKDMNRHFSKEDIYAAKGMKCCSSLAIREMOIKTT 1109
Qy 121 MRHYLTPVRMAIIKKSGNN---RDMDEAGNHSSQOTITRTKNGQTPHYLTHRW---110QS 174
Db 1110 MRHYLTPVRMAIIKKSGNNRCWRGCGEIG-----TLHGWMDCKLVOP 1152
Qy 175 HWTV---LSDISELMHKTDRIVNLL 197
Db 1153 LKSWVRFLRDL-ELFIPDPALPL 1177

```

## RESULT 10

```

012881 ID 012881 PRELIMINARY; PRT; 1275 AA.
AC 012881;

```

```

Dt 01-NOV-1996 (TREMBlrel. 01, Created)
Dt 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
Dt 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Retrotransposable L1 element LRE2 from chromosome 1q.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX TRANSPOSON=LRE2;

```

```

RA Holmes S.E., Dombroski B.A., Krebs C.M., Boehm C.D., Kazanian H.H.;
RT "A new retrotransposable human L1 element from the LRE2 locus on
RT chromosome 1q produces a chimaeric insertion.";
RL Nat. Genet. 7:143-148(1994).
DR EMBL; U09116; AAB60345.1;
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR004477; RVTse.
DR Pfam; PF03372; Exo_endo_phos. 1.
DR Pfam; PF00078; rvt. 1.
KM RNA-directed DNA polymerase.
FT VARIANT 1182 1182 E -> K.
SQ SEQUENCE 1275 AA; 149064 MW; CEFA429261FEE223 CRC64;

```

```

Query Match 39.5%; Score 672.5; DB 4; Length 1275;
Best Local Similarity 71.4%; Pred. No. 4.8e-47;
Matches 147; Conservative 4; Mismatches 28; Indels 27; Gaps 5;

```

```

Qy 1 MGDFMSKTPKAMATRAKIDKMDILKLSFCTAKETITRVNROLTEWEXIFATYSPDKGL 60
Db 990 VGKDFMSKTPKAMATRAKIDKMDILKLSFCTAKETITRVNROLTEWEXIFATYSPDKGL 1049
Qy 61 ISRIYNELKQIYKKTNNPIKKAQKDMNRHFSKEDIYAAKGMKCCSSLAIREMOIKTT 120
Db 1050 ISRIYNELKQIYKKTNNPIKKAQKDMNRHFSKEDIYAAKGMKCCSSLAIREMOIKTT 1109

```

```

QY 121 MRYHLTPVMAIIKSGNN---RDMDAGNHSQOTITRTKNQTPHYLTHRW---ILQOS 174
DB 1110 MRYHLTPVMAIIKSGNNRCRCGCEIG-----TLHCWMDCKLVOP 1152
QY 175 HWTV---LSDISELMHKTDRIVNLL 197
DB 1153 LKSWVRFLRDL-ELRIPDPALPL 1177

RESULT 11
ID 000360 PRELIMINARY; PRT; 1275 AA.
AC 000360;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative p150.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=97285120; PubMed=9140393;
RA Sassaman D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,
RA DeBerardinis R.J., Gabriel A., Swergold G.D., Kazazian H.H. Jr.;
RT "Many human LI elements are capable of retrotransposition.";
RL Nat. Genet. 16:37-43(1997).
DR EMBL; U93563; AAC51261.1;
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR Pfam; PF00078; rvt; 1.
KW RNA-directed DNA polymerase.
SQ SEQUENCE 1275 AA; 149062 MW; 350B4F0358E525F0 CRC64;

Query Match 39.5%; Score 672.5; DB 4; Length 1275;
Best Local Similarity 71.4%; Pred. No. 4.8e-47;
Matches 147; Conservative 3; Mismatches 28; Indels 27; Gaps 5;

QY 2 GQDFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQLTWEKIFATYSPDKGL 61
DB 991 GQDFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQLTWEKIFATYSPDKGL 1050
QY 62 SRIYNELKOIYKKTKNPIKAVYDMNRHPSKEGIYAAKKMKYSSSLAIRMOIKTTM 121
DB 1051 SRIYNELKOIYKKTKNPIKAVYDMNRHPSKEGIYAAKKMKYSSSLAIRMOIKTTM 1110
QY 122 RYHLTEVRAAIIKSGNN---RDMDAGNHSQOTITRTKNQTPHYLTHRW---ILQOSH 175
DB 1111 RYHLTEVRAAIIKSGNNRCRCGCEIG-----TLHCWMDCKLVOP 1153
QY 176 WTVV---LSDISELMHKTDRIVNLL 197
DB 1154 WKSVMRFLRDL-ELRIPDPALPL 1177

RESULT 12
ID 000366 PRELIMINARY; PRT; 1275 AA.
AC 000366;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative p150.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=97285120; PubMed=9140393;

```

```

RA Sassaman D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,
RA DeBerardinis R.J., Gabriel A., Swergold G.D., Kazazian H.H. Jr.;
RT "Many human LI elements are capable of retrotransposition.";
RL Nat. Genet. 16:37-43(1997).
DR EMBL; U93567; AAC51267.1;
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR Pfam; PF00078; rvt; 1.
KW RNA-directed DNA polymerase.
SQ SEQUENCE 1275 AA; 149033 MW; 07E88F8F4DB831A2 CRC64;

Query Match 39.5%; Score 672.5; DB 4; Length 1275;
Best Local Similarity 71.4%; Pred. No. 4.8e-47;
Matches 147; Conservative 4; Mismatches 28; Indels 27; Gaps 5;

QY 1 MGDFFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQLTWEKIFATYSPDKGL 60
DB 990 VGDFFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQLTWEKIFATYSPDKGL 1049
QY 61 ISRIYNELKOIYKKTKNPIKAVYDMNRHPSKEGIYAAKKMKYSSSLAIRMOIKTT 120
DB 1050 ISRIYNELKOIYKKTKNPIKAVYDMNRHPSKEGIYAAKKMKYSSSLAIRMOIKTT 1109
QY 121 MRYHLTPVMAIIKSGNN---RDMDAGNHSQOTITRTKNQTPHYLTHRW---ILQOS 174
DB 1110 MRYHLTPVMAIIKSGNNRCRCGCEIG-----TLHCWMDCKLVOP 1152
QY 175 HWTV---LSDISELMHKTDRIVNLL 197
DB 1153 LKSWVRFLRDL-ELRIPDPALPL 1177

RESULT 13
ID 000368 PRELIMINARY; PRT; 1275 AA.
AC 000368;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative p150.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=97285120; PubMed=9140393;
RA Sassaman D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,
RA DeBerardinis R.J., Gabriel A., Swergold G.D., Kazazian H.H. Jr.;
RT "Many human LI elements are capable of retrotransposition.";
RL Nat. Genet. 16:37-43(1997).
DR EMBL; U93568; AAC51269.1;
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR Pfam; PF00078; rvt; 1.
KW RNA-directed DNA polymerase.
SQ SEQUENCE 1275 AA; 149078 MW; 3BB03BD0C2E06B61 CRC64;

Query Match 39.5%; Score 672.5; DB 4; Length 1275;
Best Local Similarity 71.4%; Pred. No. 4.8e-47;
Matches 147; Conservative 4; Mismatches 28; Indels 27; Gaps 5;

QY 1 MGDFFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQLTWEKIFATYSPDKGL 60
DB 990 VGDFFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQLTWEKIFATYSPDKGL 1049
QY 61 ISRIYNELKOIYKKTKNPIKAVYDMNRHPSKEGIYAAKKMKYSSSLAIRMOIKTT 120
DB 1050 ISRIYNELKOIYKKTKNPIKAVYDMNRHPSKEGIYAAKKMKYSSSLAIRMOIKTT 1109
QY 121 MRYHLTPVMAIIKSGNN---RDMDAGNHSQOTITRTKNQTPHYLTHRW---ILQOS 174

```

```

Db 1110 MRYHLTPVRAAIKKSGNNRCWRCGCEIG-----TLHHCWMDCKLVOP 1152
Qy 175 HWVTV---LSDISELMHKTDRIVNLL 197
Db 1153 LMKSVWRFLRDL-ELRIPDPALPIL 1177

```

## RESULT 14

```

Q9UN80 PRELIMINARY; PRT; 1275 AA.
ID Q9UN80
AC Q9UN80;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Hypothetical 149.0 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON=RETROTRANSPOSON L1.
RX MEDLINE=98361157; PubMed=9697692;
RA Schwann U., Lenzner S., Dong J., Feil S., Hinzmann B.,
RA van Duijhoven G., Kirschner R., Hemberger M., Bergen A.A.,
RA Rosenberg T., Pluckers A.J., Fundele R., Rosenthal A., Cremers F.P.,
RA Ropers H.H., Berger W.,
RT "Positional cloning of the gene for X-linked retinitis pigmentosa 2.",
RL Nat. Genet. 19:327-332(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TRANSPOSON=RETROTRANSPOSON L1.
RA Kimberland M.L., Divoky V., Prechal J., Schwann U., Berger W.,
RT "Full-length human L1 insertions retain the capacity for high
RT frequency retrotransposition in cultured cells.",
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AFI48856; AAD39215.1; -
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR004777; RTase.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR Pfam; PF00078; rvc; 1.
KM Hypothetical protein; RNA-directed DNA polymerase.
SQ SEQUENCE 1275 AA; 149043 MW; A93622F87A462CE4 CRC64;

```

```

Query Match 39.5%; Score 672.5; DB 4; Length 1275;
Best Local Similarity 71.4%; Pred. No. 4.8e-47;
Matches 147; Conservative 4; Mismatches 28; Indels 27; Gaps 5;

```

```

Qy 1 MGKDFMKTTPKAMATKAKIDKMDLILKSFCTAKETIRVNRQLTEWEKIFATYSPDKGL 60
Db 990 VGKDFMKTTPKAMATKAKIDKMDLILKSFCTAKETIRVNRQLTEWEKIFATYSPDKGL 1049
Qy 61 ISRIYNEIKQIYKKKTNNPIKKWVKDNNRHFSEKGIYAAKHKYSSSLAIREMOIKTT 120
Db 1050 ISRIYNEIKQIYKKKTNNPIKKWVKDNNRHFSEKEDIYAAKHKYSSSLAIREMOIKTT 1109
Qy 121 MRYHLTPVRAAIKKSGNN---RMDPAGNHSQQTITRTKNOTPHVLTIRW---ILQOS 174
Db 1110 MRYHLTPVRAAIKKSGNNRCWRCGCEIG-----TLHHCWMDCKLVOP 1152
Qy 175 HWVTV---LSDISELMHKTDRIVNLL 197
Db 1153 LMKSVWRFLRDL-ELRIPDPALPIL 1177

```

## RESULT 15

```

ID Q15606 PRELIMINARY; PRT; 1275 AA.
AC Q15606;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)

```

```

DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Hypothetical 149.0 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Dombroski B.A.;
RL Submitted (JAN-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 84-1275 FROM N.A.
RA Woods-Samuels P.;
RL Submitted (APR-1989) to the EMBL/GenBank/DBJ databases.
DR EMBL; M80343; AAB59368.1; -
DR EMBL; M22333; AAA88037.1; -
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR004777; RTase.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR Pfam; PF00078; rvc; 1.
KM Hypothetical protein; RNA-directed DNA polymerase.
SQ SEQUENCE 1275 AA; 148971 MW; 6DA1229742848323 CRC64;

```

```

Query Match 39.5%; Score 672.5; DB 4; Length 1275;
Best Local Similarity 71.4%; Pred. No. 4.8e-47;
Matches 147; Conservative 4; Mismatches 28; Indels 27; Gaps 5;

```

```

Qy 1 MGKDFMKTTPKAMATKAKIDKMDLILKSFCTAKETIRVNRQLTEWEKIFATYSPDKGL 60
Db 990 VGKDFMKTTPKAMATKAKIDKMDLILKSFCTAKETIRVNRQLTEWEKIFATYSPDKGL 1049
Qy 61 ISRIYNEIKQIYKKKTNNPIKKWVKDNNRHFSEKGIYAAKHKYSSSLAIREMOIKTT 120
Db 1050 ISRIYNEIKQIYKKKTNNPIKKWVKDNNRHFSEKEDIYAAKHKYSSSLAIREMOIKTT 1109
Qy 121 MRYHLTPVRAAIKKSGNN---RMDPAGNHSQQTITRTKNOTPHVLTIRW---ILQOS 174
Db 1110 MRYHLTPVRAAIKKSGNNRCWRCGCEIG-----TLHHCWMDCKLVOP 1152
Qy 175 HWVTV---LSDISELMHKTDRIVNLL 197
Db 1153 LMKSVWRFLRDL-ELRIPDPALPIL 1177

```

```

Search completed: June 5, 2003, 12:43:48
Job time : 82 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: June 5, 2003, 12:39:49 ; Search time 40 Seconds  
(without alignments)  
761.865 Million cell updates/sec

Title: US-10-083-853B-1

Sequence: 1 MGDFPMSTPKAMATKAKID.....DQLDHERCDICSSRPPR 317

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 73: \*  
1: PIR1: \*  
2: PIR2: \*  
3: PIR3: \*  
4: PIR4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	679.5	39.9	1280	2 B34087	hypothetical prote
2	674	39.6	712	2 S23650	retrovirus-related
3	672.5	39.5	1275	2 I38588	reverse transcript
4	672.5	39.5	1275	2 S65824	reverse transcript
5	668.5	39.2	1275	2 B28096	line-1 protein ORF
6	658.5	38.6	1259	4 GNHUI1	retrovirus-related
7	637	37.4	370	2 JC7581	spinal cord-derive
8	585	34.3	370	2 JC7582	spinal cord-derive
9	443.5	26.0	1260	4 GNLRL1	retrovirus-related
10	425.5	25.0	275	2 S21348	probable pol polyp
11	412.5	24.2	1281	1 GNSMLL	retrovirus-related
12	374	21.9	392	2 S57662	hypothetical prote
13	344	20.2	92	2 F41925	hypothetical prote
14	343.5	20.2	562	2 JU0033	hypothetical Li pr
15	271	15.9	513	2 S21976	probable RNA-direc
16	243.5	14.3	143	2 E41925	hypothetical prote
17	125.5	7.4	148	2 D49530	16k vascular endoc
18	101	5.9	1895	2 T15881	hypothetical prote
19	100.5	5.9	625	2 H70330	hypothetical prote
20	98	5.8	4550	2 T18440	hypothetical prote
21	96.5	5.7	419	2 S69207	vascular endotheli
22	95.5	5.6	188	2 JC4680	vascular endotheli
23	95.5	5.6	207	2 JC4679	vascular endotheli
24	95	5.6	2562	2 T14266	kin protein - chic
25	94	5.5	166	2 JN0248	platelet-derived g
26	94	5.5	188	2 US0735	platelet-derived g
27	92.5	5.4	870	2 G86450	PSD1.31 protein -
28	92.5	5.4	959	2 C82392	oxidoreductase/iro
29	92.5	5.4	4092	1 S38128	dynein heavy chain

30	92	5.4	133	2 B49530	vascular endotheli
31	91.5	5.4	270	2 T47421	hypothetical prote
32	91.5	5.4	4668	2 F82885	hypothetical prote
33	91	5.3	803	2 B84931	DNA topoisomerase
34	90.5	5.3	344	2 T20209	hypothetical prote
35	90.5	5.3	357	2 B98056	hypothetical prote
36	90.5	5.3	739	2 E86434	protein F17F8.27 (
37	90.5	5.3	1817	2 D71606	hypothetical prote
38	90.5	5.3	2437	2 T18482	hypothetical prote
39	90	5.3	849	2 H84668	Mutator-like trans
40	89.5	5.3	357	2 A95190	membrane associat
41	89.5	5.3	455	2 H97237	DBP4 protein - Yea
42	89.5	5.3	704	2 S25371	hypothetical prote
43	89.5	5.3	2539	2 B71619	hypothetical prote
44	89	5.2	196	2 A37359	platelet-derived g
45	89	5.2	196	2 B28964	platelet-derived g

## ALIGNMENTS

RESULT 1  
B34087  
hypothetical protein (L1H 3' region) - human  
C/Species: Homo sapiens (man)  
C/Date: 30-Mar-1990 #sequence\_revision 30-Mar-1990 #text\_change 30-Sep-1993  
C/Accession: B34087  
R/Scott, A.F.; Schmeckpeper, B.J.; Abdelrazik, M.; Comey, C.T.; O'Hara, B.; Rossiter, J.I  
Genomics 1, 113-125, 1987  
A/Title: Origin of the human L1 elements: proposed progenitor genes deduced from a conser  
A/Reference number: A34087; PMID:88085155; PMID:3692483  
A/Accession: B34087  
A/Status: preliminary; not compared with conceptual translation  
A/Molecule type: DNA  
A/Residues: 1-1280 <SC0>  
C/Superfamily: pol polyprotein

Query Match 39.9%; Score 679.5; DB 2; Length 1280;  
Best Local Similarity 71.8%; Pred. No. 4.5e-45;  
Matches 148; Conservative 3; Mismatches 28; Indels 27; Gaps 5;

QY 1 MGDFPMSTPKAMATKAKIDKMDLILKLSFCTAKETIRVNRQLTWEKIFATYSPDKL 60  
DB 995 MGDFPMSTPKAMATKAKIDKMDLILKLSFCTAKETIRVNRQLTWEKIFATYSPDKL 1054  
QY 61 ISSIYNELKQIYKKTNPDKVVKDNRHFSKGIYAAKQNKXSSLAIREMOIKTT 120  
DB 1055 ISSIYNELKQIYKKTNPDKVVKDNRHFSKGIYAAKQNKXSSLAIREMOIKTT 1114  
QY 121 MRVHLTPVRAIIRKSGNN---RDMDEAGNHSQQTITRTKQTPHVLTRW---ILQOS 174  
DB 1115 MRVHLTPVRAIIRKSGNNRCCWGCCEIG-----TLHICWDCKLQVP 1157  
QY 175 HWTV---LSDISLNMHTDRIYNLL 197  
DB 1158 LKTVWRFLKDL-ELIIPDPRAIPL 1182

RESULT 2  
S23650  
retrovirus-related hypothetical protein II - human retrotransposon LINE-1  
C/Species: Homo sapiens (man)  
C/Date: 22-Nov-1993 #sequence\_revision 01-Nov-1996 #text\_change 08-Jan-1999  
C/Accession: S23650  
R/Hohjoh, H.; Minakami, R.; Sakaki, Y.  
Nucleic Acids Res. 18, 4093-4104, 1990  
A/Title: Selective cloning and sequence analysis of the human L1 (LINE-1) sequences which  
A/Reference number: S23649; PMID:90332398; PMID:2165587  
A/Accession: S23650  
A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-712 <HOH>  
A/Cross-references: EMBL:X52235

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1990

C/Genetics:

A/Mobile element: LINE-1

A/Start codon: GTG

C/Superfamily: pol polyprotein

Query March 39.6%; Score 674; DB 2; Length 712;  
Best Local Similarity 93.6%; Pred. No. 6.2e-45;  
Matches 131; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 MGKDFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQLTWEKIFATYSPDKGL 60  
DB 427 MGKDFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQLTWEKIFATYSPDKGL 486  
QY 61 ISRIYNELKOIYKKKTNPPIKKWVDMNRHFSKEGIYAAKKHMKYSSSLAIREMOIKTT 120  
DB 487 ISRIYNELKOIYKKKTNPPIKKWVDMNRHFSKEGIYAAKKHMKYSSSLAIREMOIKTT 546  
QY 121 MRYHLTPVRMAIIKSGNNR 140  
DB 547 MRYHLTPVRMAIIKSGNNR 566

### RESULT 3

reverse transcriptase homolog - human retrotransposon L1

N/Alternate names: ORF2 protein

C/Species: Homo sapiens (man)

C/Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 26-Aug-1999

C/Accession: I38588

A/Title: A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q F

A/Reference number: I38587; MUID:95004577; PMID:7920631

A/Accession: I38588

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1275 <RES>

A/Cross-references: EMBL:U09116; NID:9483914; PIDN:AA860345.1; PID:9483916

C/Superfamily: pol polyprotein

Query Match 39.5%; Score 672.5; DB 2; Length 1275;  
Best Local Similarity 71.4%; Pred. No. 1.6e-44;  
Matches 147; Conservative 4; Mismatches 28; Indels 27; Gaps 5;

QY 1 MGKDFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQLTWEKIFATYSPDKGL 60  
DB 990 VGKDFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQLTWEKIFATYSPDKGL 1049  
QY 61 ISRIYNELKOIYKKKTNPPIKKWVDMNRHFSKEGIYAAKKHMKYSSSLAIREMOIKTT 120  
DB 1050 ISRIYNELKOIYKKKTNPPIKKWVDMNRHFSKEGIYAAKKHMKYSSSLAIREMOIKTT 1109  
QY 121 MRYHLTPVRMAIIKSGNNR--RDMDAGNHHSQQTITRTKNOTPHVLTTRW---IIQOS 174  
DB 1110 MRYHLTPVRMAIIKSGNNRCMRGCGEIG-----TLHGMWDCKLVOP 1152  
QY 175 HWYTV---ISDISELMHKTDRIVNLL 197  
DB 1153 LMKSVWRFRLDL-ELRIFPDPAIPLL 1177

### RESULT 4

reverse transcriptase homolog - human transposon L1.1

C/Species: Homo sapiens (man)

C/Date: 20-Jul-1996 #sequence\_revision 13-Mar-1997 #text\_change 26-Aug-1999

C/Accession: S65824

R/Dombroski, B.A.

submitted to the EMBL Data Library, January 1992

A/Description: Isolation of an active human transposable element.

A/Reference number: S65823

A/Accession: S65824

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-1275 <DOM>

A/Cross-references: EMBL:M00340; NID:G339767; PIDN:AAA51622.1; PID:G339771

C/Superfamily: pol polyprotein

Query Match 39.5%; Score 672.5; DB 2; Length 1275;  
Best Local Similarity 71.4%; Pred. No. 1.6e-44;  
Matches 147; Conservative 4; Mismatches 28; Indels 27; Gaps 5;

QY 1 MGKDFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQLTWEKIFATYSPDKGL 60  
DB 990 VGKDFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQLTWEKIFATYSPDKGL 1049  
QY 61 ISRIYNELKOIYKKKTNPPIKKWVDMNRHFSKEGIYAAKKHMKYSSSLAIREMOIKTT 120  
DB 1050 ISRIYNELKOIYKKKTNPPIKKWVDMNRHFSKEGIYAAKKHMKYSSSLAIREMOIKTT 1109  
QY 121 MRYHLTPVRMAIIKSGNNR--RDMDAGNHHSQQTITRTKNOTPHVLTTRW---IIQOS 174  
DB 1110 MRYHLTPVRMAIIKSGNNRCMRGCGEIG-----TLHGMWDCKLVOP 1152  
QY 175 HWYTV---ISDISELMHKTDRIVNLL 197  
DB 1153 LMKSVWRFRLDL-ELRIFPDPAIPLL 1177

### RESULT 5

line-1 protein ORF2 - human

C/Species: Homo sapiens (man)

C/Date: 03-Nov-1988 #sequence\_revision 03-Nov-1988 #text\_change 30-Sep-1993

C/Accession: B28096

A/Title: Unit-length line-1 transcripts in human teratocarcinoma cells.

A/Reference number: A28096; MUID:88246405; PMID:2454389

A/Accession: B28096

A/Status: preliminary; not compared with conceptual translation

A/Molecule type: mRNA

A/Residues: 1-1275 <SKO>

C/Superfamily: pol polyprotein

Query Match 39.2%; Score 668.5; DB 2; Length 1275;  
Best Local Similarity 70.9%; Pred. No. 3.3e-44;  
Matches 146; Conservative 4; Mismatches 29; Indels 27; Gaps 5;

QY 1 MGKDFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQLTWEKIFATYSPDKGL 60  
DB 990 VGKDFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQLTWEKIFATYSPDKGL 1049  
QY 61 ISRIYNELKOIYKKKTNPPIKKWVDMNRHFSKEGIYAAKKHMKYSSSLAIREMOIKTT 120  
DB 1050 ISRIYNELKOIYKKKTNPPIKKWVDMNRHFSKEGIYAAKKHMKYSSSLAIREMOIKTT 1109  
QY 121 MRYHLTPVRMAIIKSGNNR--RDMDAGNHHSQQTITRTKNOTPHVLTTRW---IIQOS 174  
DB 1110 MRYHLTPVRMAIIKSGNNRCMRGCGEIG-----TLHGMWDCKLVOP 1152  
QY 175 HWYTV---ISDISELMHKTDRIVNLL 197  
DB 1153 LMKSVWRFRLDL-ELRIFPDPAIPLL 1177

### RESULT 6

retrovirus-related reverse transcriptase pseudogene - human

C/Species: Homo sapiens (man)

C/Date: 31-Mar-1988 #sequence\_revision 04-Jan-1996 #text\_change 14-May-1999

C/Accession: A25313

R/Hatori, M.; Kunata, S.; Takenaka, O.; Sakaki, Y.

Nature 321, 625-628, 1986

A/Title: L1 family of repetitive DNA sequences in primates may be derived from a sequence



A:Reference number: A93381; MUID:86230917; PMID:2423883  
A:Accession: A2513  
A>Status: conceptual translation of pseudogene  
A:Molecule type: DNA  
A:Residues: 1-1259 <HAT>  
A>Note: this sequence was constructed from an alignment of published and unpublished seq  
C:Keywords: reverse transcriptase; pseudogene

Query Match 38.6%; Score 658.5; DB 4; Length 1259;  
Best Local Similarity 69.9%; Pred. No. 1.9e-43;

Matches 144; Conservative 6; Mismatches 29; Indels 27; Gaps 5;

QY 1 MGDFMSKTPKAMATKADIKMDILKLSFCTAKETIRVNRQLTWEMKIFATYSPDKL 60  
DB MGDFMTPKPKAMATKADIKMDILKLSFCTAKETIRVNRQLTWEMKIFATYSPDKL 1048

QY 61 ISRIYVELKQIYKKTNPPIKKVVKDMNRHFSKEGIYAANKKKYSSSLAIREMOIKTT 120  
DB ISRIYVELKQIYKKTNPPIKKVVKDMNRHFSKEGIYAANKKKYSSSLAIREMOIKTT 1108

QY 121 MRHILTPVPMALIKKSGNN---RDMDEAGNHSSOQITTRKNQTPHVLTHRW---ILQOS 174  
DB MRHILTPVPMALIKKSGNNRCCGCGEIG-----TLHGMWCKLQVOP 1151

QY 175 HWVTV---LSDISELMHKTRIVNLL 197  
DB HWVTV---LSDISELMHKTRIVNLL 1176

QY 1152 LMKSVMRFLRL-ELRIFPDPAIPPL 1176  
DB 1152 LMKSVMRFLRL-ELRIFPDPAIPPL 1176

## RESULT 7

spinal cord-derived growth factor-B precursor - human

C:Species: Homo sapiens (man)  
C:Date: 30-Jun-2001 #sequence\_revision 30-Jun-2001 #text\_change 24-Aug-2001

C:Accession: J07591

R:Hamada, T.; Uti-Tel, K.; Imaki, J.; Miyata, Y.

Biochem. Biophys. Res. Commun. 280, 733-737, 2001

A>Title: Molecular cloning of SCDF-B, a novel growth factor homologous to SCDF/PDGF-C/

F;272-370/Region: homologous to platelet-derived growth factor/vascular endothelial grow

A:Reference number: J07591; MUID:21092670; PMID:11162582

A:Accession: J07591

A:Molecule type: DNA

A:Residues: 1-370 <HAM>

A:Cross-references: DDBJ:AB033832

C:Genetics:

A:Gene: scdgb-B

F;1-117/Domain: secretory signal sequence #status predicted <SIG>

F;18-370/Product: spinal cord-derived growth factor-B #status predicted <MAT>

F;52-170/Region: CUB domain #status predicted

F;272-370/Region: homologous to platelet-derived growth factor/vascular endothelial grow

F;294-308/Region: conserved motif #status predicted

Query Match 37.4%; Score 637; DB 2; Length 370;  
Best Local Similarity 65.4%; Pred. No. 2.3e-42;

Matches 125; Conservative 9; Mismatches 11; Indels 46; Gaps 3;

QY 173 QSHWVTVLSDISELMHKT-----DRIVNLMC-----M 200  
DB QSHWVTVLSDISELMHKT-----DRIVNLMC-----M 239

QY 201 YLLT-----VDLRLNDAKRYSCTPRNYSVNIREEELKLANVVEPRCLAV 246  
DB YLLT-----VDLRLNDAKRYSCTPRNYSVNIREEELKLANVVEPRCLAV 239

QY 247 QRCGNGCGCTVNRSGTCSGKTVKHYEVLQFEPGHIRGRGAKTALVDIOLDHHER 306  
DB QRCGNGCGCTVNRSGTCSGKTVKHYEVLQFEPGHIRGRGAKTALVDIOLDHHER 359

QY 307 CDCICSSRPR 317  
DB CDCICSSRPR 370

## RESULT 8

J07592  
spinal cord-derived growth factor-B precursor - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 30-Jun-2001 #sequence\_revision 30-Jun-2001 #text\_change 24-Aug-2001

C:Accession: J07592

R:Hamada, T.; Uti-Tel, K.; Imaki, J.; Miyata, Y.

Biochem. Biophys. Res. Commun. 280, 733-737, 2001

A>Title: Molecular cloning of SCDF-B, a novel growth factor homologous to SCDF/PDGF-C/

F;272-370/Region: homologous to platelet-derived growth factor/vascular endothelial grow

A:Reference number: J07591; MUID:21092670; PMID:11162582

A:Accession: J07592

A:Molecule type: mRNA

A:Residues: 1-370 <HAM>

A:Cross-references: DDBJ:AB052170

C:Genetics:

A:Gene: scdgb-B

F;1-117/Domain: secretory signal sequence #status predicted <SIG>

F;18-370/Product: spinal cord-derived growth factor-B #status predicted <MAT>

F;52-170/Region: CUB domain #status predicted

F;272-370/Region: homologous to platelet-derived growth factor/vascular endothelial grow

F;294-308/Region: conserved motif #status predicted

Query Match 34.3%; Score 585; DB 2; Length 370;  
Best Local Similarity 89.4%; Pred. No. 2.7e-38;

Matches 101; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 205 VDLRLNDAKRYSCTPRNYSVNIREEELKLANVVEPRCLAVORCGNGCGCTVNRSGT 264  
DB VDLRLNDAKRYSCTPRNYSVNIREEELKLANVVEPRCLAVORCGNGCGCTVNRSGT 317

QY 265 CNSGKTVKHYEVLQFEPGHIRGRGAKTALVDIOLDHHERCDICSSRPR 317  
DB CNSGKTVKHYEVLQFEPGHIRGRGAKTALVDIOLDHHERCDICSSRPR 370

## RESULT 9

GNRL1  
retrovirus-related reverse transcriptase pseudogene - slow loris

C:Species: Nycticebus coucang (slow loris)

C:Date: 31-Mar-1988 #sequence\_revision 04-Jan-1996 #text\_change 14-May-1999

C:Accession: B25313

R:Hattori, M.; Kuhara, S.; Takenaka, O.; Sakaki, Y.

Nature 321, 625-628, 1986

A>Title: LI family of repetitive DNA sequences in primates may be derived from a sequence

A:Reference number: A93381; MUID:86230917; PMID:2423883

A:Accession: B25313

A>Status: conceptual translation of pseudogene

A:Molecule type: DNA

A:Residues: 1-1260 <HAT>

A>Note: this sequence was constructed from an alignment of six sequences, determined by t

C:Keywords: reverse transcriptase; pseudogene

Query Match 26.0%; Score 443.5; DB 4; Length 1260;  
Best Local Similarity 50.3%; Pred. No. 1.3e-26;

Matches 95; Conservative 25; Mismatches 52; Indels 17; Gaps 2;

QY 1 MGDFMSKTPKAMATKADIKMDILKLSFCTAKETIRVNRQLTWEMKIFATYSPDKL 60  
DB MGDFMSKTPKAMATKADIKMDILKLSFCTAKETIRVNRQLTWEMKIFATYSPDKL 1048

QY 61 ISRIYVELKQIYKKTNPPIKKVVKDMNRHFSKEGIYAANKKKYSSSLAIREMOIKTT 120  
DB ISRIYVELKQIYKKTNPPIKKVVKDMNRHFSKEGIYAANKKKYSSSLAIREMOIKTT 1108

QY 121 MRHILTPVPMALIKKSGNNRDMDEAGNHSSOQITTRKNQTPHVLTHRW---ILQSHAV 177  
DB MRHILTPVPMALIKKSGNNRDMDEAGNHSSOQITTRKNQTPHVLTHRW---ILQSHAV 1154

QY 178 TVLSDISEL 186  
DB TVLSDISEL 1163



A:Cross-references: GB:M69297; NID:g182218; PIDN:AA58463.1; PID:g182220  
A>Note: sequence extracted from NCBI backbone (NCBI:P:106603)  
C:superfamily: pol polyprotein

Query Match 20.2%; Score 344; DB 2; Length 92;  
Best Local Similarity 82.4%; Pred. No. 4.1e-20;  
Matches 70; Conservative 1; Mismatches 4; Indels 10; Gaps 1;

QY 1 MGKDFMSKTPKAMATKAKIDKMDLIKLSKSC-----TKETTRVNRQLTMEKI 50  
DB 1 MGKDFMSKTPKAMATKAKIDKMDLIKLSKSCSLGKTKENTYAKETTRVNRQLTMEKI 60

QY 51 FATYSPDKGLISRIYNELKQIYKKK 75  
DB 61 FATYSSDKGLISRIYNELKQIYKKK 85

## RESULT 14

JU0033

hypothetical L1 protein (third intron of gene TS) - human

C:Species: Homo sapiens (man)

C&gt;Date: 07-Jun-1990 #sequence\_revision 07-Jun-1990 #text\_change 13-Sep-1996

C:Accession: JU0033

R:Horie, N.; Nalbandov, J.; Kaneda, S.; Ayusawa, D.; Seno, T.; Takeishi, K.

J. Biochem. 106, 1-4, 1989

A&gt;Title: Identification and characterization of an L1 family sequence with a very long C

A:Reference number: JU0033; NMID:89380111; PMID:2476429

A:Accession: JU0033

A&gt;Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-562 &lt;HOR&gt;

A&gt;Note: this sequence is similar to human teratocarcinoma L1 RNA species and RNA depende

Query Match 20.2%; Score 343.5; DB 2; Length 562;  
Best Local Similarity 90.8%; Pred. No. 3.5e-19;  
Matches 69; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

QY 1 MGKDFMSKTPKAMATKAKIDKMDLIKLSKSCETKETTIRVNRQLTMEKIFATYSPDKGL 60  
DB 479 MGKDFM-EMPKAMATKAKIDKMDLIKLSKSCETKETTIRVNRQLTMEKIFATYSPDKGL 537

QY 61 ISRIYNELKQIYKKKT 76  
DB 538 ISRIYNELKQIYKKKT 553

## RESULT 15

S21976

probable RNA-directed DNA polymerase (EC 2.7.7.49) (clone MH2C) - rat retrotransposon L1

N:Alternate names: reverse transcriptase

C:Species: Rattus norvegicus (Norway rat)

C&gt;Date: 19-Mar-1997 #sequence\_revision 17-Oct-1997 #text\_change 26-Aug-1999

C:Accession: S21976

R:Khare, O.; Iives, H.; Speck, M.

submitted to the EMBL Data Library, August 1991

A:Reference number: S16783

A:Accession: S21976

A:Molecule type: mRNA

A:Residues: 1-513 &lt;RAH&gt;

A:Cross-references: EMBL:X61295; NID:G56521; PIDN:CAA43593.1; PID:G56522

A:Experimental source: clone MH2C

C:Genetics:

A:Mobile element: retrotransposon L1

C:superfamily: pol polyprotein

C:Keywords: nucleotidyltransferase; polyprotein; reverse transcriptase

Query Match 15.9%; Score 271; DB 2; Length 513;  
Best Local Similarity 59.3%; Pred. No. 1.5e-13;  
Matches 48; Conservative 18; Mismatches 15; Indels 0; Gaps 0;

QY 2 GKDFMSKTPKAMATKAKIDKMDLIKLSKSCETKETTIRVNRQLTMEKIFATYSPDKGL 61  
DB 2 GKDFMSKTPKAMATKAKIDKMDLIKLSKSCETKETTIRVNRQLTMEKIFATYSPDKGL 61

DB 433 GKDFMSKTPKAMATKAKIDKMDLIKLSKSCETKETTIRVNRQLTMEKIFATYSPDKGL 492

QY 62 SRIYNELKQIYKKKTGNPIKK 82

DB 493 SKIYELKGLDRRETNNPIKK 513

Search completed: June 5, 2003, 12:44:35  
Job time: 42 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2003, 12:32:13 ; Search time 70 Seconds

(without alignments)  
603.435 Million cell updates/sec

Title: US-10-083-853B-1

Perfect score: 1704  
Sequence: 1 MGKDFMSKTPKAWATAKID.....DIQLDHERDCICSRPPR 317

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_101002:.\*  
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	818.5	48.0	939	22	ABG12508
2	818.5	48.0	939	22	ABG14439
3	818.5	48.0	940	22	ABG07054
4	818.5	48.0	736	22	ABG04462
5	812.5	47.7	969	22	ABG14779
6	812.5	47.7	1074	22	ABG10255
7	804.5	47.2	1074	22	ABG07890
8	804	47.1	482	22	ABG17719
9	803	47.1	853	22	ABG14765
10	783.5	46.0	1680	22	ABG17645

11	783.5	46.0	1825	22	ABG11906	Novel human diagno
12	779	45.7	279	22	ABG15456	Novel human diagno
13	771	45.2	827	22	AA130932	Novel human secreto
14	760.5	44.6	1067	22	ABG18495	Novel human diagno
15	758.5	44.5	1356	22	ABG10381	Novel human diagno
16	758.5	44.5	1721	22	ABG08687	Novel human diagno
17	757.5	44.5	1655	22	ABG14955	Novel human diagno
18	755.5	44.3	566	22	ABG07412	Novel human diagno
19	755.5	44.3	536	22	ABG12326	Novel human diagno
20	755.5	44.3	1361	22	ABG23198	Novel human diagno
21	755.5	44.3	2764	22	ABG10129	Novel human diagno
22	755	44.3	1261	22	ABG07064	Novel human diagno
23	744.5	43.7	1565	22	ABG06288	Novel human diagno
24	744	43.7	1669	22	ABG13314	Novel human diagno
25	736	43.2	894	22	ABG08978	Novel human diagno
26	736	43.2	894	22	ABG13059	Novel human diagno
27	730	42.8	662	22	ABG05640	Novel human diagno
28	730	42.8	662	22	ABG07422	Novel human diagno
29	730	42.8	662	22	ABG14256	Novel human diagno
30	728	42.7	944	22	ABG02301	Novel human diagno
31	728	42.7	644	22	ABG18485	Novel human diagno
32	727	42.7	671	22	ABG17890	Novel human diagno
33	726	42.6	908	22	ABG05638	Novel human diagno
34	726	42.6	908	22	ABG12511	Novel human diagno
35	725	42.5	597	22	ABG07850	Novel human diagno
36	725	42.5	597	22	ABG10065	Novel human diagno
37	725	42.5	703	22	ABG14431	Novel human diagno
38	725	42.5	703	22	ABG23675	Novel human diagno
39	725	42.5	754	22	ABG12695	Novel human diagno
40	724	42.5	452	22	ABG06567	Novel human diagno
41	724	42.5	959	22	ABG02299	Novel human diagno
42	720	42.3	2678	22	ABG11124	Novel human diagno
43	718.5	42.2	1344	22	ABG13058	Novel human diagno
44	718	42.1	775	22	ABG02472	Novel human diagno
45	715	42.0	1078	22	ABG17206	Novel human diagno

#### ALIGNMENTS

RESULT 1	ABG12508	standard; Protein; 939 AA.
ID	ABG12508	
AC	ABG12508	
DT	18-FEB-2002	(first entry)
XX		
XX		Novel human diagnostic protein #12499.
DE		
XX		
XX		Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW		Food supplement; medical imaging; diagnostic; genetic disorder.
OS		Homo sapiens.
XX		
PN	WO200175067-A2.	
PD	11-OCT-2001.	
XX		
PF	30-MAR-2001; 2001WO-US08631.	
XX		
PR	31-MAR-2000; 2000US-0540217.	
PR	23-AUG-2000; 2000US-0649167.	
XX		
PA	(HYSE-) HYSEQ INC.	
XX		
PI	Dmanac RT, Liu C, Tang YF,	
XX		
DR	WPI; 2001-639362/73.	
XX		
XX	N-PADB; AAST6695.	
PT		New isolated polynucleotide and encoded polypeptides, useful in
		diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID No 42867; 103bp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 939 AA;  
Query Match 48.0%; Score 818.5; DB 22; Length 939;  
Best Local Similarity 82.4%; Pred. No. 2.1e-75;  
Matches 159; Conservative 9; Mismatches 18; Indels 7; Gaps 1;  
QY 1 MGKPMSTPKAMATKADIKMDLIKSFCTAKETTRVNRQUTEMEKIFATYSPDKGL 60  
DB 685 MGKPMSTPKAMATKADIKMDLIKSFCTAKETTRVNRQUTEMEKIFATYSPDKGL 744  
QY 61 ISRIYNELKQIYKKTKNPIKKNVKNRHRFSKGIYAAKGMKKYSSSLAIREMOIKTT 120  
DB 745 ISRIHNEKQIYKKTKNPNKTKWAKDMNRHFSKEDIYAAKGMKKCSSSLAIREMOIKTT 804  
QY 121 MRYHLTPVMAIIRKSGNNRDMDEAGNHSQOITTRTKNQTPHVLTHRWL-----QQ 173  
DB 805 MRYHLTPVMAVIRKSGNNRDMDEAGNHSQOITVTRTKNQTPHVLTHRWLNNETWTQE 864  
QY 174 SHWTVLSDISEL 186  
DB 865 GEYHTLGTVISEL 877  
RESULT 2  
ABG14439  
ID ABG14439 standard; Protein; 939 AA.  
XX  
AC ABG14439;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #14430.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.

XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS78626.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID No 44798; 103bp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 939 AA;  
Query Match 48.0%; Score 818.5; DB 22; Length 939;  
Best Local Similarity 82.4%; Pred. No. 2.1e-75;  
Matches 159; Conservative 9; Mismatches 18; Indels 7; Gaps 1;  
QY 1 MGKPMSTPKAMATKADIKMDLIKSFCTAKETTRVNRQUTEMEKIFATYSPDKGL 60  
DB 685 MGKPMSTPKAMATKADIKMDLIKSFCTAKETTRVNRQUTEMEKIFATYSPDKGL 744  
QY 61 ISRIYNELKQIYKKTKNPIKKNVKNRHRFSKGIYAAKGMKKYSSSLAIREMOIKTT 120  
DB 745 ISRIHNEKQIYKKTKNPNKTKWAKDMNRHFSKEDIYAAKGMKKCSSSLAIREMOIKTT 804  
QY 121 MRYHLTPVMAIIRKSGNNRDMDEAGNHSQOITTRTKNQTPHVLTHRWL-----QQ 173  
DB 805 MRYHLTPVMAVIRKSGNNRDMDEAGNHSQOITVTRTKNQTPHVLTHRWLNNETWTQE 864  
QY 174 SHWTVLSDISEL 186  
DB 865 GEYHTLGTVISEL 877  
RESULT 3  
ABG07054  
ID ABG07054 standard; Protein; 940 AA.  
XX  
AC ABG07054;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #7045.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX

OS Homo sapiens.  
 XX MO200175067-A2.  
 XX 11-OCT-2001.  
 PD 30-MAR-2001; 2001WO-US08631.  
 PF 31-MAR-2000; 2000US-0540217.  
 XX PR 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 XX Drmanac RT, Liu C, Tang YT,  
 PI N-PSDB; AAS71241.  
 DR WPI; 2001-639362/73.  
 XX N-PSDB; AAS71241.  
 DR  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensic, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 PS Claim 20; SEQ ID No 37413; 103bp; English.  
 XX  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensic, gene mapping, identification of mutations in  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 XX Sequence 940 AA;  
 SQ  
 Query Match 48.0%; Score 818.5; DB 22; Length 940;  
 Best Local Similarity 82.4%; Pred. No. 2.1e-75;  
 Matches 159; Conservative 9; Mismatches 18; Indels 7; Gaps 1;  
 OY 1 MGKDFMSKTPKAMATKAKIDKMDLILKLSFCTAKETTRIVNRQLTWEMKIFATYSPDKGL 60  
 DB 686 MGKDFMSKTPKAMATKAKIDKMDLILKLSFCTAKETTRIVNRQLTWEMKIFATYSPDKGL 745  
 OY 61 ISRIYNELKQIYKKTKNPIKQVVKDMNRHFSKEGIYAAKKMKYSSSLAIREMOIKTT 120  
 DB 746 ISRIYNELKQIYKKTKNPIKQVVKDMNRHFSKEGIYAAKKMKYSSSLAIREMOIKTT 805  
 OY 121 MRYHLTPVMAIILKSGNNRDMDEAGNHSQOTITRTKQTPHVLTHRWIL-----QQ 173  
 DB 806 MRYHLTPVMAIILKSGNNRDMDEAGNHSQOTITRTKQTPHVLTHRWELNNEITWTQE 865  
 OY 174 SHWVTVLSDISEL 186  
 DB 866 GEYHTLGTVTSSEL 878  
 RESULT 4  
 ABG04462  
 ID ABG04462 standard; Protein; 736 AA.

XX ABG04462;  
 AC 13-FEB-2002 (first entry)  
 XX  
 DT Novel human diagnostic protein #4453.  
 XX  
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX food supplement; medical imaging; diagnostic; genetic disorder.  
 XX Homo sapiens.  
 OS  
 XX MO200175067-A2.  
 XX 11-OCT-2001.  
 PD 30-MAR-2001; 2001WO-US08631.  
 PF 31-MAR-2000; 2000US-0540217.  
 XX PR 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 XX Drmanac RT, Liu C, Tang YT,  
 PI N-PSDB; AAS68649.  
 DR WPI; 2001-639362/73.  
 XX N-PSDB; AAS68649.  
 DR  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensic, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 PS Claim 20; SEQ ID No 34821; 103bp; English.  
 XX  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensic, gene mapping, identification of mutations in  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 XX Sequence 736 AA;  
 SQ  
 Query Match 48.0%; Score 818; DB 22; Length 736;  
 Best Local Similarity 82.1%; Pred. No. 1.7e-75;  
 Matches 160; Conservative 7; Mismatches 18; Indels 10; Gaps 2;  
 OY 1 MGKDFMSKTPKAMATKAKIDKMDLILKLSFCTAKETTRIVNRQLTWEMKIFATYSPDKGL 60  
 DB 176 MGKDFMSKTPKAMATKAKIDKMDLILKLSFCTAKETTRIVNRQLTWEMKIFATYSPDKGL 235  
 OY 61 ISRIYNELKQIYKKTKNPIKQVVKDMNRHFSKEGIYAAKKMKYSSSLAIREMOIKTT 120  
 DB 236 ISRIYNELKQIYKKTKNPIKQVVKDMNRHFSKEGIYAAKKMKYSSSLAIREMOIKTT 295  
 OY 121 MRYHLTPVMAIILKSGNNRDMDEAGNHSQOTITRTKQTPHVLTHRWIL--QOSHVV-- 177

Db 296 MRYHLTPVRMAIIKSGNNRDMDEGNHSGQOTIARTKQNTLPHVLTIRHRELNNENTWQOE 355  
Qy 178 -----TVLSDISP 185  
Db 356 GEHHTLETFVGDMDDE 370

## RESULT 5

ABG14779  
ID ABG14779 standard; Protein; 969 AA.  
AC ABG14779;  
XX  
XX 18-FEB-2002 (first entry)  
DT  
DE Novel human diagnostic protein #14770.  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX Homo sapiens.  
OS  
PN WO200175067-A2.  
PD 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
PF  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX  
XX Drmanac RT, Liu C, Tang YT;  
PI  
XX  
XX WPI; 2001-639362/73.  
DR  
DR N-PSDB; AAS78966.  
XX

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
XX  
XX Claim 20; SEQ ID No 45138; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SO Sequence 969 AA;

Query Match 47.7%; Score 812.5; DB 22; Length 969;  
Best Local Similarity 88.7%; Fred. No. 9.1e-75;  
Matches 157; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 1 MGKDFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRIVNROUTEMEKIFATYSPDKGL 60  
Db 779 MGKDFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRIVNROPTTEKEKIFATYSSPKGL 838  
Qy 61 ISRIYNELKQIYKKTKNPIKKWVKDMNRHPSKEGIYAARKKKYSSSLAIREMOIKTT 120  
Db 839 ISRIYNELKQIYKKTKNPIKKWVKDMNRHPSKEDIYAAARKKKCSSSLAIREMOIKTT 898  
Qy 121 MRYHLTPVRMAIIKSGNNRDMDEGNHSGQOTIARTKQNTLPHVLTIRHRL-00SHW 176  
Db 899 MRYHLTPVRMAIIKSGNNRDMDEGNHSGQOTIARTKQNTLPHVLTIRHRELNNENTW 955

## RESULT 6

ABG10255  
ID ABG10255 standard; Protein; 1074 AA.  
AC ABG10255;  
XX  
XX 13-FEB-2002 (first entry)  
DT  
DE Novel human diagnostic protein #10246.  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX Homo sapiens.  
OS  
PN WO200175067-A2.  
PD 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
PF  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX  
XX Drmanac RT, Liu C, Tang YT;  
PI  
XX  
XX WPI; 2001-639362/73.  
DR  
DR N-PSDB; AAS74442.  
XX

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
XX  
XX Claim 20; SEQ ID No 40614; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX



SQ Sequence 1074 AA;  
 Query Match 47.7%; Score 812.5; DB 22; Length 1074;  
 Best Local Similarity 82.9%; Pred. No. 1.1e-74;  
 Matches 160; Conservative 8; Mismatches 18; Indels 7; Gaps 1;

QY 1 MGKDFMSKTPKAMATKAKIDKMDLKLKSFCTAKETIRVNROLTWEMKIFATYSPDKGL 60  
 DB MGKDFMSKTPKAMATKAKIDKMDLKLKSFCTAKETIRVNROLTWEMKIFATYSPDKGL 879

QY 61 ISRIYNELKQIYKKTKNPIKKVQKDMNRHPSKEGIYAANKMKYSSSLAIRMOIKTT 120  
 DB ISRIYNELKQIYKKTKNPIKKVQKDMNRHPSKEGIYAANKMKYSSSLAIRMOIKTT 939

QY 121 MRYHLTPVPMALIKKSGNNRDMDEAGNHSOQITRTKQTPHVLTHRWTL-----QQ 173  
 DB 940 VRHLLTPVPMALIKKSGNNRDMDEAGNHSOQITRTKQTPHVLTHRWTLNEITWTQ 999

QY 174 SHWTVLSDISEL 186  
 DB 1000 GEYHTLTGTVTSEL 1012

RESULT 7  
 ABG07890  
 ID ABG07890 standard; Protein; 1074 AA.  
 XX ABG07890;  
 AC  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #7881.  
 XX  
 KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KM Food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 PA (HYSE-) HYSEQ INC.  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI, 2001-639362/73.  
 DR N-PSDB; AAS72077.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20; SEQ ID No 36249; 103bp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (II) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 1074 AA;  
 Query Match 47.2%; Score 804.5; DB 22; Length 1074;  
 Best Local Similarity 81.9%; Pred. No. 7.1e-74;  
 Matches 158; Conservative 9; Mismatches 19; Indels 7; Gaps 1;

QY 1 MGKDFMSKTPKAMATKAKIDKMDLKLKSFCTAKETIRVNROLTWEMKIFATYSPDKGL 60  
 DB MGKDFMSKTPKAMATKAKIDKMDLKLKSFCTAKETIRVNROLTWEMKIFATYSPDKGL 879

QY 61 ISRIYNELKQIYKKTKNPIKKVQKDMNRHPSKEGIYAANKMKYSSSLAIRMOIKTT 120  
 DB ISRIYNELKQIYKKTKNPIKKVQKDMNRHPSKEGIYAANKMKYSSSLAIRMOIKTT 939

QY 121 MRYHLTPVPMALIKKSGNNRDMDEAGNHSOQITRTKQTPHVLTHRWTL-----QQ 173  
 DB 940 MRYHLTPVPMALIKKSGNNRDMDEAGNHSOQITRTKQTPHVLTHRWTLNEITWTQ 999

QY 174 SHWTVLSDISEL 186  
 DB 1000 GEYHTLTGTVTSEL 1012

RESULT 8  
 ABG17719  
 ID ABG17719 standard; Protein; 482 AA.  
 XX ABG17719;  
 AC  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #17710.  
 XX  
 KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KM Food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 PA (HYSE-) HYSEQ INC.  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI, 2001-639362/73.  
 DR N-PSDB; AAS81906.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20; SEQ ID No 48078; 103bp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SO Sequence 482 AA;

Query Match 47.2%; Score 804; DB 22; Length 482;  
Best Local Similarity 91.1%; Pred. No. 2.6e-74;  
Matches 154; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 1 MGKDFMSKTPKAMATKAKIDKMDLILKSFCTAKETTRVNRDLTEWEKIFATYSPDKGL 60  
DB 237 MGKDFMSKTPKAMATKAKIDKMDLILKSFCTAKETTRVNRDLTEWEKIFATYSPDKGL 296  
QY 61 ISRIYNELKOIYKKKTKNPIKKWVDMNRHFSKEGIYAAKGMKKYSSSLAIREMOIKTT 120  
DB 297 ISRIYNELKOIYKKKTKNPIKKWVDMNRHFSKEDIYAAKGMKKYSSSLAIREMOIKTT 356  
QY 121 MRYHLTPVRMAIIKKSGNNRDMDEAGNHSQOITTRTKNQTPIHLTH 169  
DB 357 MRYHLTPVRMAIIKKSGNNRDMDEAGNHSQOITTRTKNQTPIHLTH 405

RESULT 9  
ABG14765  
ID ABG14765 standard; Protein: 853 AA.  
XX  
AC ABG14765;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #14756.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN MO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
XX  
N-PSDB; AAS76952.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess

PT biodiversity  
XX  
XX Claim 20; SEQ ID No 45124; 103pp; English.  
XX

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probe,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SO Sequence 853 AA;

Query Match 47.1%; Score 803; DB 22; Length 853;  
Best Local Similarity 91.0%; Pred. No. 7.3e-74;  
Matches 152; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 1 MGKDFMSKTPKAMATKAKIDKMDLILKSFCTAKETTRVNRDLTEWEKIFATYSPDKGL 60  
DB 686 MGKDFMSKTPKAMATKAKIDKMDLILKSFCTAKETTRVNRDLTEWEKIFATYSPDKGL 745  
QY 61 ISRIYNELKOIYKKKTKNPIKKWVDMNRHFSKEGIYAAKGMKKYSSSLAIREMOIKTT 120  
DB 746 ISRIYNELKOIYKKKTKNPIKKWVDMNRHFSKEDIYAAKGMKKYSSSLAIREMOIKTT 805  
QY 121 MRYHLTPVRMAIIKKSGNNRDMDEAGNHSQOITTRTKNQTPIHLTH 167  
DB 806 MRYHLTPVRMAIIKKSGNNRDMDEAGNHSQOITTRTKNQTPIHLTH 852

RESULT 10  
ABG17645  
ID ABG17645 standard; Protein: 1680 AA.  
XX  
AC ABG17645;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #17636.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN MO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
XX  
N-PSDB; AAS76952.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess

DR MPI; 2001-639362/73.  
DR N-PSDB; AAS81832.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID No 48004; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1680 AA;  
Query Match 46.0%; Score 783.5; DB 22; Length 1680;  
Best Local Similarity 86.9%; Pred. No. 2e-71;  
Matches 153; Conservative 5; Mismatches 17; Indels 1; Gaps 1;  
QY 1 MGKDFMSKTPKAMATYAKIDKMDLILKSFCTAKETTRIVNRQLTEWEKIFATYSDKGL 60  
DB 990 MGKDFISKTPKAMATYAKIDKMDLILKSFCTAKETTRIVNRQPIEMEKIFATYSSDKGL 1049  
QY 61 ISRIYVELKQIYKKKTKN-PIKKVDMNRPHEKGIYAKKMKYSSSLAREMOIKT 119  
DB 1050 ISRIYVELKQIYKKKTKNPIKKVDMNRPHEKGIYAKKMKYSSSLAREMOIKT 1109  
QY 120 TMRVHLTPVRMAIIKSGNNRDMDEAGNHSQOTITRTKNQTPHVLTHRMILQOSH 175  
DB 1110 TMRVHLTPVRMAIIKSGNNRDMDEAGNHSQOTITRTKNQTPHVLTHRMILQOSH 1165  
RESULT 11  
ABG11906  
ID ABG11906 standard; Protein; 1825 AA.  
XX  
AC ABG11906;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #11897.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YH;  
XX  
XX MPI; 2001-639362/73.  
DR N-PSDB; AAS76093.  
XX  
PS Claim 20; SEQ ID No 42265; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1825 AA;  
Query Match 46.0%; Score 783.5; DB 22; Length 1825;  
Best Local Similarity 85.9%; Pred. No. 2.3e-71;  
Matches 152; Conservative 5; Mismatches 19; Indels 1; Gaps 1;  
QY 1 MGKDFMSKTPKAMATYAKIDKMDLILKSFCTAKETTRIVNRQLTEWEKIFATYSDKGL 60  
DB 426 MGKDFMSKTPKAMATYAKIDKMDLILKSFCTAKETTRIVNRQPIEMEKIFATYSSDKGR 485  
QY 61 ISRIYVELKQIYKKKTKNPIKKVDMNRPHEKGIYAKKMKYSSSLAREMOIKT 120  
DB 486 ISRIYVELKQIYKKKTKNPIKKVDMNRPHEKGIYAKKMKYSSSLAREMOIKT 545  
QY 121 MRVHLTPVRMAIIKSGNNRDMDEAGNHSQOTITRTKNQTPHVLTHRMILQOSH-W 176  
DB 546 MRVHLTPVRMAIIKSGNNRDMDEAGNHSQOTITRTKNQTPHVLTHRMILQOSH-W 602  
RESULT 12  
ABG15456  
ID ABG15456 standard; Protein; 279 AA.  
XX  
AC ABG15456;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #15447.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.

XX 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001MO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
XX  
XX 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
XX  
XX N-PSDB; AAS79643.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity.  
XX  
XX Claim 20; SEQ ID No 45815; 103bp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. ABG00010-ABG30377 represent novel human  
XX diagnostic amino acid sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 279 AA;  
SQ  
Query Match 45.7%; Score 779; DB 22; Length 279;  
Best Local Similarity 88.6%; Pred. No. 4.6e-72;  
Matches 148; Conservative 5; Mismatches 14; Indels 0; Gaps 0;  
QY 1 MGKDFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQLTEWEKIFATYSPDKGL 60  
DB 1 MGKDFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQLTEWEKIFATYSPDKGL 60  
QY 61 ISRIYNELKOIYKKTKNPIKRVKVMNRHPSKEGIYAAGKMKYSSSLAIREMOIKTT 120  
DB 61 ISRIYNELKOIYKKTKNPIKRVKVMNRHPSKEGIYAAGKMKYSSSLAIREMOIKTT 120  
QY 61 ISRIYNELKOIYKKTKNPIKRVKVMNRHPSKEGIYAAGKMKYSSSLAIREMOIKTT 120  
DB 61 ISRIYNELKOIYKKTKNPIKRVKVMNRHPSKEGIYAAGKMKYSSSLAIREMOIKTT 120  
QY 121 MRHYLTPVRRAIIKKSNNRDMDEAGNHSOQITTRTGNQTPHYLTH 167  
DB 121 MRHYLTPVRRAIIKKSNNRDMDEAGNHSOQITTRTGNQTPHYLTH 167  
QY 121 MRHYLTPVRRAIIKKSNNRDMDEAGNHSOQITTRTGNQTPHYLTH 167  
DB 121 MRHYLTPVRRAIIKKSNNRDMDEAGNHSOQITTRTGNQTPHYLTH 167  
RESULT 13  
AAU30932  
ID AAU30932 standard; Protein; 827 AA.  
XX  
XX AAU30932;  
XX  
XX 18-DEC-2001 (first entry)  
XX  
XX Novel human secreted protein #1423.  
XX

KW Human; vaccination; gene therapy; nutritional supplement;  
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;  
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.  
XX  
XX Homo sapiens.  
XX  
XX WO200179449-A2.  
XX  
XX 25-OCT-2001.  
XX  
XX 16-APR-2001; 2001MO-US08656.  
XX  
XX 18-APR-2000; 2000US-0552929.  
XX  
XX 26-JAN-2001; 2001US-0770160.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-611725/70.  
XX  
XX Nucleic acids encoding a range of human polypeptides, useful in genetic  
XX vaccination, testing and therapy -  
XX  
XX Claim 20; Page 378; 765pp; English.  
XX  
XX The invention relates to novel human secreted polypeptides. The  
XX polypeptides and antibodies to the polypeptides are useful for  
XX determining the presence of or predisposition to a disease associated  
XX with altered levels of polypeptide. The polypeptides are also useful for  
XX identifying agents (agonists and antagonists) that bind to them. Cells  
XX expressing the proteins are useful for identifying a therapeutic agent  
XX for use in treatment of a pathology related to aberrant expression or  
XX physiological interactions of the polypeptide. Vectors comprising  
XX the nucleic acids encoding the polypeptides and cells genetically  
XX engineered to express them are also useful for producing the proteins.  
XX The proteins are useful in genetic vaccination, testing and  
XX therapy, and can be used as nutritional supplements. They may be used to  
XX increase stem cell proliferation; to regulate haematopoiesis; and in  
XX bone, cartilage, tendon and/or nerve tissue growth or regeneration;  
XX immune suppression and/or stimulation; as anti-inflammatory agents; and  
XX in treatment of leukaemias. AAU29510-AAU3304 represent the amino acid  
XX sequences of novel human secreted proteins of the invention.  
XX  
XX Sequence 827 AA;  
SQ  
Query Match 45.2%; Score 771; DB 22; Length 827;  
Best Local Similarity 76.3%; Pred. No. 1.4e-70;  
Matches 151; Conservative 12; Mismatches 29; Indels 6; Gaps 2;  
QY 1 MGKDFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQLTEWEKIFATYSPDKGL 60  
DB 556 MGKDFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQLTEWEKIFATYSPDKGL 615  
QY 61 ISRIYNELKOIYKKTKNPIKRVKVMNRHPSKEGIYAAGKMKYSSSLAIREMOIKTT 120  
DB 61 ISRIYNELKOIYKKTKNPIKRVKVMNRHPSKEGIYAAGKMKYSSSLAIREMOIKTT 120  
QY 616 ISRIYNELKOIYKKTKNPIKRVKVMNRHPSKEGIYAAGKMKYSSSLAIREMOIKTT 675  
DB 616 ISRIYNELKOIYKKTKNPIKRVKVMNRHPSKEGIYAAGKMKYSSSLAIREMOIKTT 675  
QY 121 MRHYLTPVRRAIIKKSNNRDMDEAGNHSOQITTRTGNQTPHYLTH 179  
DB 676 IRHYLTPVRRAIIKKSNNRDMDEAGNHSOQITTRTGNQTPHYLTH 179  
QY 180 LSDISELMKTKTDRIYNLL 197  
DB 736 -----EEHHTLGPVIGVI 748  
RESULT 14  
ABG18495  
ID ABG18495 standard; Protein; 1067 AA.  
XX  
XX ABG18495;  
XX



Qy 61 ISRIYNELKOIYKKKTNP I KKWVKDNNRHPSEKGIYA AKKMKYSSLAIREMOIKTT 120  
Db 727 ISRIYNELKOIYKKKTNP I KKWVKDNNRHPSEKDIYA AKKMKYSSLAIREMOIKTT 786  
Qy 121 MRYHLTPVRMAIIKKSNNRMDDEAGNHSSOOTITRTKNOTPHVLTHERWILQOSH-W 176  
Db 787 MRYHLTPVRMAIIKKSNNRMDGAGGHYPRQTNEGTENQIPDVLTCMELNDKNIW 843

Search completed: June 5, 2003, 12:41:51  
Job time : 72 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2003, 12:40:39 ; Search time 26 Seconds  
(without alignments)  
358.733 Million cell updates/sec

Title: US-10-083-853B-1

Sequence: 1 MGDPMFKPMKAMATKAKID.....DIQDHERDCICSSRPPR 317

Scoring table: BLOSUM62

Gapop 10.0 , Gapept 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	637	37.4	370	US-09-457-066-37	Sequence 37, Appl
2	637	37.4	370	US-09-540-224-2	Sequence 2, Appl
3	588	34.5	370	US-09-540-224-4	Sequence 4, Appl
4	318.5	18.7	345	US-09-040-220-2	Sequence 2, Appl
5	318.5	18.7	345	US-09-457-066-2	Sequence 2, Appl
6	318.5	18.7	345	US-09-265-686-2	Sequence 2, Appl
7	318.5	18.7	345	US-09-540-224-5	Sequence 5, Appl
8	314.5	18.5	345	US-09-457-066-43	Sequence 43, Appl
9	118	6.9	24	US-09-540-224-9	Sequence 9, Appl
10	114.5	6.7	354	US-08-915-795-5	Sequence 5, Appl
11	112.5	6.6	325	US-08-915-795-3	Sequence 3, Appl
12	106.5	6.2	321	US-08-915-795-9	Sequence 9, Appl
13	106.5	6.2	358	US-08-915-795-8	Sequence 8, Appl
14	102	6.0	109	US-09-469-186-1	Sequence 1, Appl
15	96.5	5.7	350	US-08-999-811-4	Sequence 4, Appl
16	96.5	5.7	350	US-08-824-996-2	Sequence 2, Appl
17	96.5	5.7	350	US-09-042-105-4	Sequence 4, Appl
18	96.5	5.7	350	US-08-510-133A-33	Sequence 33, Appl
19	96.5	5.7	350	US-08-885-895-33	Sequence 33, Appl
20	96.5	5.7	419	US-08-999-811-2	Sequence 2, Appl
21	96.5	5.7	419	US-09-042-105-2	Sequence 2, Appl
22	96.5	5.7	419	US-09-042-105-18	Sequence 18, Appl
23	96.5	5.7	419	US-08-795-430-8	Sequence 8, Appl
24	96.5	5.7	419	US-08-510-133A-35	Sequence 35, Appl
25	96.5	5.7	419	US-09-355-700-8	Sequence 8, Appl
26	96.5	5.7	419	US-08-601-132-33	Sequence 33, Appl
27	96.5	5.7	419	PCT-US96-09001-2	Sequence 2, Appl

28	95.5	5.6	102	1	US-08-469-427A-2	Sequence 2, Appl
29	95.5	5.6	102	2	US-08-609-443B-2	Sequence 2, Appl
30	95.5	5.6	102	2	US-08-569-063C-2	Sequence 2, Appl
31	95.5	5.6	102	4	US-08-851-896-2	Sequence 2, Appl
32	95.5	5.6	133	1	US-08-469-427A-9	Sequence 9, Appl
33	95.5	5.6	133	2	US-08-609-443B-9	Sequence 9, Appl
34	95.5	5.6	133	2	US-08-569-063C-9	Sequence 9, Appl
35	95.5	5.6	133	4	US-08-851-896-9	Sequence 9, Appl
36	95.5	5.6	188	1	US-08-469-427A-5	Sequence 5, Appl
37	95.5	5.6	188	2	US-08-609-443B-5	Sequence 5, Appl
38	95.5	5.6	188	2	US-08-569-063C-5	Sequence 5, Appl
39	95.5	5.6	188	4	US-08-851-896-5	Sequence 5, Appl
40	95.5	5.6	207	2	US-08-609-443B-13	Sequence 13, Appl
41	95.5	5.6	207	2	US-08-569-063C-13	Sequence 13, Appl
42	95.5	5.6	207	2	US-08-851-896-13	Sequence 13, Appl
43	94	5.5	195	1	US-08-469-427A-7	Sequence 7, Appl
44	94	5.5	195	2	US-08-609-443B-7	Sequence 7, Appl
45	94	5.5	195	2	US-08-569-063C-7	Sequence 7, Appl

## ALIGNMENTS

```
RESULT 1
US-09-457-066-37
; Sequence 37, Application US/09457066
; Patent No. 6432673
; GENERAL INFORMATION:
; APPLICANT: Gao, Zeren
; APPLICANT: Hart, Charles E.
; APPLICANT: Piddington, Christopher S.
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Shoemaker, Kimberly E.
; APPLICANT: Gilbertson, Debra G.
; APPLICANT: West, James W.
; TITLE OF INVENTION: GROWTH FACTOR HOMOLOG ZVEGF3
; FILE REFERENCE: 98-60
; CURRENT APPLICATION NUMBER: US/09/457,066
; CURRENT FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 37
; LENGTH: 370
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-457-066-37

Query Match
Best Local Similarity 37.4%; Score 637; DB 4; Length 370;
Matches 125; Conservative 9; Mismatches 11; Indels 46; Gaps 3;

QY 173 QSHWTVLSDISLHMKT-----DRIVNLMC-----M 200
Db 180 ETIMESVTSISGVSYNSPVTPTLIALDPAKIAEFTVEDLTKYFNPESQOELENN 239
QY 201 YLT-----VDLRINDAKRYSCPTPNYSVNIIEELKIANVVEPPCLIV 246
Db 240 YLTPTPRGSGYDRKSKVDLRINDAKRYSCPTPNYSVNIIEELKIANVVEPPCLIV 299
QY 247 QRCGNCGGCTVWNRSCCTNSGKTVKKYHEVLQFEFGHKKRGRATMALVDIQDHER 306
Db 300 QRCGNCGGCTVWNRSCCTNSGKTVKKYHEVLQFEFGHKKRGRATMALVDIQDHER 359
QY 307 CDCICSSRPPR 317
Db 360 CDCICSSRPPR 370

RESULT 2
US-09-540-224-2
; Sequence 2, Application US/09540224
; Patent No. 6468543
; GENERAL INFORMATION:
```

APPLICANT: Gilbertson, Debra G.  
APPLICANT: Hart, Charles E.  
TITLE OF INVENTION: METHODS FOR PROMOTING GROWTH OF BONE,  
TITLE OF INVENTION: LIGAMENT AND CARTILAGE USING ZVEGF4  
FILE REFERENCE: 00-28  
CURRENT APPLICATION NUMBER: US/09/540,224  
CURRENT FILING DATE: 2000-03-31  
EARLIER APPLICATION NUMBER: US 60/180,169  
EARLIER FILING DATE: 2000-02-04  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 370  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-540-224-2

Query Match 37.4%; Score 637; DB 4; Length 370;  
Best Local Similarity 65.4%; Pred. No. 5,8e-59;  
Matches 125; Conservative 9; Mismatches 11; Indels 46; Gaps 3;

QY 173 QSHWTVLSIDSELMHKT-----DRVNLIMC-----M 200  
DB 180 ETNWSYVSISGVSNPSYTDPTLIADALDKIAEPDVEDLKYENPESWOEDLENN 239  
QY 201 YLLT-----VDLRLNDADAKRYSCPTPNYSVNIREEKLKLANVVFPRCLLV 246  
DB 240 YLDPRIRKGRSYHDKRSKYVDLRLNDADAKRYSCPTPNYSVNIREEKLKLANVVFPRCLLV 299  
QY 247 QRCGNGCGGTVMNRSCCTNSGKTVKXVHEVLOFEPGHIKRGRAKTMALVDIOLDHHER 306  
DB 300 QRCGNGCGGTVMNRSCCTNSGKTVKXVHEVLOFEPGHIKRGRAKTMALVDIOLDHHER 359  
QY 307 CDCICSSRPPR 317  
DB 360 CDCICSSRPPR 370

RESULT 3  
US-09-540-224-4  
Sequence 4, Application US/09540224  
Patent No. 6468543  
GENERAL INFORMATION:  
APPLICANT: Gilbertson, Debra G.  
APPLICANT: Hart, Charles E.  
TITLE OF INVENTION: METHODS FOR PROMOTING GROWTH OF BONE,  
TITLE OF INVENTION: LIGAMENT AND CARTILAGE USING ZVEGF4  
FILE REFERENCE: 00-28  
CURRENT APPLICATION NUMBER: US/09/540,224  
CURRENT FILING DATE: 2000-03-31  
EARLIER APPLICATION NUMBER: US 60/180,169  
EARLIER FILING DATE: 2000-02-04  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 4  
LENGTH: 370  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-540-224-4

Query Match 34.5%; Score 588; DB 4; Length 370;  
Best Local Similarity 90.3%; Pred. No. 8,6e-54;  
Matches 102; Conservative 6; Mismatches 5; Indels 0; Gaps 0;  
QY 205 VDLRLNDADAKRYSCPTPNYSVNIREEKLKLANVVFPRCLLVORCGNGCGGTVMNRSC 264  
DB 258 VDLRLNDADAKRYSCPTPNYSVNIREEKLKLANVVFPRCLLVORCGNGCGGTVMNRSC 317  
QY 265 CNSGKTVKXVHEVLOFEPGHIKRGRAKTMALVDIOLDHHERCCTICSSRPPR 317  
DB 318 CNSGKTVKXVHEVLOFEPGHIKRGRAKTMALVDIOLDHHERCCTICSSRPPR 370

RESULT 4  
US-09-040-220D-2  
Sequence 2, Application US/09040220D  
Patent No. 691311  
GENERAL INFORMATION:  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Kuo, Sophia S.  
TITLE OF INVENTION: NOVEL POLYPEPTIDES HAVING HOMOLOGY TO VASCULAR  
TITLE OF INVENTION: ENDOTHELIAL CELL GROWTH FACTOR AND BONE MORPHOGENETIC  
TITLE OF INVENTION: PROTEIN 1 AND NUCLEIC ACIDS ENCODING SAME, THEIR USES,  
FILE REFERENCE: P1122  
CURRENT APPLICATION NUMBER: US/09/040,220D  
CURRENT FILING DATE: 1998-03-17  
NUMBER OF SEQ ID NOS: 8  
SEQ ID NO 2  
LENGTH: 345  
TYPE: PRT  
ORGANISM: Human  
US-09-040-220D-2

Query Match 18.7%; Score 318.5; DB 4; Length 345;  
Best Local Similarity 53.3%; Pred. No. 2,2e-25;  
Matches 57; Conservative 14; Mismatches 33; Indels 3; Gaps 1;

QY 205 VDLRLNDADAKRYSCPTPNYSVNIREEKLKLANVVFPRCLLVORCGNGCGGTVMNRSC 264  
DB 236 VDLRLNLEEVRLISCTPNYSVNIREEKLKRTDTTFWEGCLLVKRGNGACCLHNCBECQ 295  
QY 265 CNSGKTVKXVHEVLOFEPGHIKRGRAKTMALVDIOLDHHERCCTIC 311  
DB 296 CVPSKTVKXVHEVQLRP---KTGVRLHKSLLTDVALHHEBECQVC 339

RESULT 5  
US-09-457-066-2  
Sequence 2, Application US/09457066  
Patent No. 6432673  
GENERAL INFORMATION:  
APPLICANT: Gao, Zeren  
APPLICANT: Hart, Charles E.  
APPLICANT: Piddington, Christopher S.  
APPLICANT: Sheppard, Paul O.  
APPLICANT: Shoemaker, Kimberly E.  
APPLICANT: Gilbertson, Debra G.  
TITLE OF INVENTION: GROWTH FACTOR HOMOLOGY ZVEGF3  
FILE REFERENCE: 98-60  
CURRENT APPLICATION NUMBER: US/09/457,066  
CURRENT FILING DATE: 1999-12-07  
NUMBER OF SEQ ID NOS: 50  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 345  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-457-066-2

Query Match 18.7%; Score 318.5; DB 4; Length 345;  
Best Local Similarity 53.3%; Pred. No. 2,2e-25;  
Matches 57; Conservative 14; Mismatches 33; Indels 3; Gaps 1;

QY 205 VDLRLNDADAKRYSCPTPNYSVNIREEKLKLANVVFPRCLLVORCGNGCGGTVMNRSC 264  
DB 236 VDLRLNLEEVRLISCTPNYSVNIREEKLKRTDTTFWEGCLLVKRGNGACCLHNCBECQ 295  
QY 265 CNSGKTVKXVHEVLOFEPGHIKRGRAKTMALVDIOLDHHERCCTIC 311  
DB 296 CVPSKTVKXVHEVQLRP---KTGVRLHKSLLTDVALHHEBECQVC 339

RESULT 6  
US-09-265-686-2



```

: Sequence 2, Application US/09265686
: Patent No. 6455283
: GENERAL INFORMATION:
: APPLICANT: Ferrara, Napoleone
: APPLICANT: Kuo, Sophia S.
: TITLE OF INVENTION: POLYPEPTIDES HOMOLOGOUS TO VEGF AND BMP1
: FILE REFERENCE: P1122P2
: CURRENT APPLICATION NUMBER: US/09/265,686
: CURRENT FILING DATE: 1999-03-10
: PRIOR APPLICATION NUMBER: US 09/040,220
: PRIOR FILING DATE: 1998-03-17
: PRIOR APPLICATION NUMBER: US 09/184,216
: PRIOR FILING DATE: 1998-11-02
: NUMBER OF SEQ ID NOS: 8
: SEQ ID NO 2
: LENGTH: 345
: TYPE: PRT
: ORGANISM: Human
: US-09-265-686-2

```

	Query Match	Similarity	18.7%	Score 318.5	DB 4	Length 345
Best Local	53.3%	Pred. No. 2.2e-25				
Matches	57	Conservative	14	Mismatches	33	Indels 3; Gaps 1
Qy	205	VDLRLINDAR	KYSTPRNYSNITREELKIANVPEFPICLYVORCGSGNGCGTVMNRGT	264		
Db	236	VDLNLITREVALYSTPRNFSVSTIEELKRDITITWPGCLVYKRCGNCACCLHNCBCQ	295			
Qy	265	CNSGKTVKKTIEVLTQFEPGHIKRRGRATMALVDIQLDHHKRCDCIC	311			
Db	296	CVPSKVTYKTYHIVLQVR--KTGVGRGLAKSLTDVALEHHEKDCVC	339			

```

RESULT 7
US-09-540-224-5
: Sequence 5, Application US/09540224
: Patent No. 6468543
: GENERAL INFORMATION:
: APPLICANT: Gilbertson, Debra G.
: APPLICANT: Hart, Charles B.
: TITLE OF INVENTION: METHODS FOR PROMOTING GROWTH OF BONE,
: TITLE OF INVENTION: LIGAMENT AND CARTILAGE USING ZVEGF4
: FILE REFERENCE: 00-28
: CURRENT APPLICATION NUMBER: US/09/540,224
: CURRENT FILING DATE: 2000-03-31
: EARLIER APPLICATION NUMBER: US 60/180,169
: EARLIER FILING DATE: 2000-02-04
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 5
: LENGTH: 345
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-540-224-5

```

	Query Match	Similarity	Score	DB 4	length
Best Local	53.3%;	Pred. No. 2.2e-25;			
Matches	57;	Conservative	14;	Mismatches	33;
				Indels	3;
				Gaps	1

```

RESULT 8
US-09-457-066-43
; Sequence 43, Application US/09457066
; Patent No. 6432673
; GENERAL INFORMATION:

```

```

1  APPLICANT: Gao, Zeren
2  APPLICANT: Hart, Charles E.
3  APPLICANT: Piddington, Christopher S.
4  APPLICANT: Sheppard, Paul O.
5  APPLICANT: Shoemaker, Kimberly E.
6  APPLICANT: Gilbertson, Debra G.
7  APPLICANT: West, James W.
8  TITLE OF INVENTION: GROWTH FACTOR HOMOLOG ZVEGF3
9  FILE REFERENCE: 98-60
10 CURRENT APPLICATION NUMBER: US/09/457,066
11 CURRENT FILING DATE: 1999-12-07
12 NUMBER OF SEQ ID NOS: 50
13 SOFTWARE: FastSeq for Windows Version 3.0
14 SEQ ID NO 43
15 LENGTH: 345
16 TYPE: PR1
17 ORGANISM: Mus musculus
18 US-09-457-066-43

```

	Query Match	18.5%;	Score 314.5;	DB 4;	Length 345;
	Best Local Similarity	43.1%;	Pred. No. 5.7e-25;		
	Matches	66;	Conservative	19;	Mismatches 47; Indels 21; Gaps 3,
OY	VTVLSDISELMH--KTDRIYNLMCKMTLLT-----			VDDLRLINDAKRYS	218
	: : : : :			: : : :	
Db	190 VTAFSTEEELIRLYDEPDRQVDDLSLTKPTQLGLGKAFLYGKKSKRYVALINTLKKEEVKLVS				249
OY	CTPNFNSVINIREELKANVFPEPCLLVORGGNGCGTVNWRSCCTCNSGTIVKKYHEVL				278
	:   :   :   :   :   :   :   :				
Db	250 CTPNFNSVISIREELKRDTJTFWPGCLLVKRGNGCACCLHNCCQCVPRTVKTKKYHEVL				309
OY	QFEFGHIKRRGRATKMVLVDIOLDHHRCDIC				311
	:				
Db	310 QLRP---KTGYGLHKSLVDVALEHHBCDVC				339

```

RESULT 9
US-09-540-224-9
; Sequence 9, Application US/09540224
; Patent No. 6468543
; GENERAL INFORMATION:
; APPLICANT: Hart, Charles E. Debra G.
; TITLE OF INVENTION: METHODS FOR PROMOTING GROWTH OF BONE,
; FILE OF INVENTION: LIGAMENT AND CARTILAGE USING ZWGF4
; FILE REFERENCE: 00-28
; CURRENT APPLICATION NUMBER: US/09/540,224
; CURRENT FILING DATE: 2000-03-31
; EARLIER APPLICATION NUMBER: US 60/180,169
; EARLIER FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 24
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: peptide
US-09-540-224-9

```

	Query Match	6.9%	Score 118;	DJ 4;	Length 24;
	Best Local Similarity	100.0%;	Pred. No.	6.9e-06;	
	Matches 23; Conservative	0;	Mismatches	0;	Indels
Gy	280 FEPGHKRRGAKTMAVDIOLD	302			
Db	2 FEPGHKRRGAKTMAVDIOLD	24			

RESULT 10  
US-08-915-795-5  
; Sequence 5, Application US/08915795  
; Patent No. 6235713

GENERAL INFORMATION:  
APPLICANT: Marc G. ACHEN  
APPLICANT: Andrew F. WILKS  
APPLICANT: Steven A. STACKER  
APPLICANT: Kari ALITALO  
TITLE OF INVENTION: GROWTH FACTOR  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Evenson, McKeown, Edwards & Lenahan P.L.L.C.  
STREET: 1200 G Street, NW, Suite 700  
CITY: Washington  
STATE: DC  
COUNTRY: United States of America  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/915,795  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: EVANS, Joseph D.  
REGISTRATION NUMBER: 26,269  
REFERENCE/DOCKET NUMBER: 1064/42983  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-8800  
TELEFAX: (202) 628-8844  
TELEX: N/A  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 354 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: NO  
ORIGINAL SOURCE:  
TISSUE TYPE: Human Lung  
US-08-915-795-5

Query Match 6.7%; Score 114.5; DB 4; Length 354;  
Best Local Similarity 24.5%; Pred. No. 0.00076;  
Matches 47; Conservative 27; Mismatches 71; Indels 47; Gaps 9;

QY 145 AGNH-----SQOTTRKQTPHVLTHRNILQOSH-----WTVL-----SDISE 185  
DB 22 SSNHEGPVKRSSQSTLERSEQIRAAASLEELRITHSEDKLWRCRLKLSFTSMDSRS 81  
QY 186 LHMHTDRIVNLMCMYLLTVLDRLNDAKRYSCTPRNVSYNIEEL-KLANVFFPRCL 244  
DB 82 ASHRSRTPATFYDIETLV-----IDEMWORTQSPRETCVAVSELGKSTNTPFKPCVC 137  
QY 245 LVORCGNCGCGTVNWSCTGNSGKT---VKYHEV---LOFEPGHIKRGRATMALVD 298  
DB 138 NVFPGCGGC-----NESELICMTSTSYISKQLEISVPLTSV-----ELVP 180  
QY 299 IQLDHERDCI 310  
DB 181 VKVANHTGCKCL 192

RESULT 11  
US-08-915-795-3  
Sequence 3, Application US/08915795  
Patent No. 6235713  
GENERAL INFORMATION:  
APPLICANT: Marc G. ACHEN  
APPLICANT: Andrew F. WILKS  
APPLICANT: Steven A. STACKER  
APPLICANT: Kari ALITALO

TITLE OF INVENTION: GROWTH FACTOR  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Evenson, McKeown, Edwards & Lenahan P.L.L.C.  
STREET: 1200 G Street, NW, Suite 700  
CITY: Washington  
STATE: DC  
COUNTRY: United States of America  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/915,795  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: EVANS, Joseph D.  
REGISTRATION NUMBER: 26,269  
REFERENCE/DOCKET NUMBER: 1064/42983  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-8800  
TELEFAX: (202) 628-8844  
TELEX: N/A  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 325 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: NO  
ORIGINAL SOURCE:  
TISSUE TYPE: Human Breast  
US-08-915-795-3

Query Match 6.6%; Score 112.5; DB 4; Length 325;  
Best Local Similarity 24.9%; Pred. No. 0.0011;  
Matches 45; Conservative 26; Mismatches 69; Indels 41; Gaps 8;

QY 150 SQOTTRKQTPHVLTHRNILQOSH-----WTVL-----SDISELMHRTDRIVNL 196  
DB 4 SQSTLERSEQIRAAASLEELRITHSEDKLWRCRLKLSFTSMDSRSASHSTFAAT 63  
QY 197 LMCYLLTVLDRLNDAKRYSCTPRNVSYNIEEL-KLANVFFPRCLLVORCGNCGC 255  
DB 64 FYDIETLV-----IDEMWORTQSPRETCVAVSELGKSTNTPFKPCVCNVFPGCGGC-- 117  
QY 256 GTVNWSCGNSGKT---VKYHEV---LOFEPGHIKRGRATMALVDIQLDHERDCI 309  
DB 118 ---NESELICMTSTSYISKQLEISVPLTSV-----ELVPKVANHTGCKC 162  
QY 310 I 310  
DB 163 L 163

RESULT 12  
US-08-915-795-9  
Sequence 9, Application US/08915795  
Patent No. 6235713  
GENERAL INFORMATION:  
APPLICANT: Marc G. ACHEN  
APPLICANT: Andrew F. WILKS  
APPLICANT: Steven A. STACKER  
APPLICANT: Kari ALITALO  
TITLE OF INVENTION: GROWTH FACTOR  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Evenson, McKeown, Edwards & Lenahan P.L.L.C.  
STREET: 1200 G Street, NW, Suite 700



Oy 210 LINDAKRYSCPTPRNYSVIREEL--KLANVFEPRCLIVQRCGNGCGTVMNRCTCNSG 268  
 Db 10 IDEEMQRTQCPSETCEVASELKGSTNFFKPCVNVFRCGGCC---NEESLICMNT 64  
 Oy 269 KT---VKYHEV---LOFEPGHKRGKAKTMAVLDIOLDHERCDCI 310  
 Db 65 STYISNQLFEISVPLTSPV-----ELVPVAVNHTGCKL 100

RESULT 15

US-08-999-811-4  
 ; Sequence 4, Application US/08999811  
 ; Patent No. 5932540  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HU, JING-SHAN  
 ; APPLICANT: ROSEN, CRAIG A.  
 ; APPLICANT: CAO, LIANG  
 ; TITLE OF INVENTION: VASCULAR ENDOTHELIAL GROWTH FACTOR 2  
 ; NUMBER OF SEQUENCES: 15  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX  
 ; STREET: 1100 NEW YORK AVENUE  
 ; CITY: WASHINGTON  
 ; STATE: DC  
 ; COUNTRY: USA  
 ; ZIP: 20005  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/999,811  
 ; FILING DATE: HERewith  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/207,550  
 ; FILING DATE: 8-MAR-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/465,968  
 ; FILING DATE: 06-JUN-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: MARKOWICZ, KAREN R.  
 ; REGISTRATION NUMBER: 36,351  
 ; REFERENCE/DOCKET NUMBER: 1488.1000004  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202)371-2600  
 ; TELEFAX: (202)371-2540  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 350 amino acids  
 ; TYPE: amino acid  
 ; \* TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-999-811-4

Query Match 5 7\*; Score 96.5; DB 2; Length 350;

Best Local Similarity 19.8%; Pred. No. 0.06; Mismatches 64; Indels 77; Gaps 9;

Matches 41; Conservative 25; Mismatches 64; Indels 77; Gaps 9;  
 Oy 184 SELAKTKDRIYNLMCMWLLVDDLRINDAKRYSCPTPRNYSVIREELKLANVVF--PR 242  
 Db 28 ANLMSRTEETIKFAAAHYNTETI-LKSIDNEMKTKQCMREVCIDVGKEFGVATNTFFKPP 86  
 Oy 243 CLVQRCGCGNG-----CGTVMNRCTCNSGK 269  
 Db 87 CVSVYRCGGCGNSBGLQCMNTSTSYLSKTLFEITVPLSQPKPVTISFANHTSCRCMSKL 146  
 Oy 270 TV-KYHEVLO-----FEPGHK-----RGRKTM 295  
 Db 147 DVYRQVSHIIRSLPATLPQCAANKTCPTNYMMNHHICRLAQEDFMFSSDAGDSDTG 206

Oy 296 LVDI-----OLDHERCDICSS--RP 315  
 Db 207 FHDICGNKEID--ETQCVCYCRAGLR 232

Search completed: June 5, 2003, 12:45:08  
 Job time : 27 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2003, 12:43:54 ; Search time 46 Seconds  
(Without alignments)  
711.460 Million cell updates/sec

Title: US-10-083-853B-1

Sequence: 1 MGDFPMKTPKAMATYAKID.....DIQLDHERDCICSSRPPR 317

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 392085 seqs, 103240269 residues

Total number of hits satisfying chosen parameters: 392085

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubppa/US08\_NEW\_PUB pep:\*
- 2: /cgn2\_6/ptodata/1/pubppa/PTCT\_NEW\_PUB pep:\*
- 3: /cgn2\_6/ptodata/1/pubppa/US06\_NEW\_PUB pep:\*
- 4: /cgn2\_6/ptodata/1/pubppa/US07\_NEW\_PUB pep:\*
- 5: /cgn2\_6/ptodata/1/pubppa/US07\_PUBCOMB pep:\*
- 6: /cgn2\_6/ptodata/1/pubppa/US07\_PUBCOMB pep:\*
- 7: /cgn2\_6/ptodata/1/pubppa/PTCTUS\_PUBCOMB pep:\*
- 8: /cgn2\_6/ptodata/1/pubppa/US08\_PUBCOMB pep:\*
- 9: /cgn2\_6/ptodata/1/pubppa/US09\_NEW\_PUB pep:\*
- 10: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB pep:\*
- 11: /cgn2\_6/ptodata/1/pubppa/US10\_NEW\_PUB pep:\*
- 12: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB pep:\*
- 13: /cgn2\_6/ptodata/1/pubppa/US60\_NEW\_PUB pep:\*
- 14: /cgn2\_6/ptodata/1/pubppa/US60\_PUBCOMB pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1704	100.0	317	US-10-083-853-2	Sequence 2, Appl1
2	680.5	39.9	956	US-10-000-2564-153	Sequence 153, App
3	672.5	39.5	917	US-10-001-835-189	Sequence 189, App
4	672.5	39.5	1275	US-10-025-201-3	Sequence 3, Appl1
5	637	37.4	322	US-10-086-623-6	Sequence 6, Appl1
6	637	37.4	322	US-10-260-539-6	Sequence 186, App
7	637	37.4	364	US-10-028-072-186	Sequence 186, App
8	637	37.4	364	US-10-131-049-186	Sequence 186, App
9	637	37.4	364	US-10-123-904-186	Sequence 186, App
10	637	37.4	364	US-10-140-470-186	Sequence 186, App
11	637	37.4	364	US-10-175-746-186	Sequence 186, App
12	637	37.4	364	US-10-176-918-186	Sequence 186, App
13	637	37.4	364	US-10-137-865-186	Sequence 186, App
14	637	37.4	364	US-10-137-865-186	Sequence 186, App
15	637	37.4	364	US-10-140-474-186	Sequence 186, App
16	637	37.4	364	US-10-142-431-186	Sequence 186, App
17	637	37.4	364	US-10-143-114-186	Sequence 186, App
18	637	37.4	364	US-10-140-002-186	Sequence 186, App
19	637	37.4	364	US-10-142-419-186	Sequence 186, App

20	637	37.4	364	9	US-10-123-262-186	Sequence 186, App
21	637	37.4	364	9	US-10-142-423-186	Sequence 186, App
22	637	37.4	364	9	US-10-121-050-186	Sequence 186, App
23	637	37.4	364	9	US-10-141-755-186	Sequence 186, App
24	637	37.4	364	9	US-10-143-032-186	Sequence 186, App
25	637	37.4	364	9	US-10-123-108-186	Sequence 186, App
26	637	37.4	364	9	US-10-123-236-186	Sequence 186, App
27	637	37.4	364	9	US-10-123-261-186	Sequence 186, App
28	637	37.4	364	9	US-10-140-921-186	Sequence 186, App
29	637	37.4	364	9	US-10-140-928-186	Sequence 186, App
30	637	37.4	364	9	US-10-121-045-186	Sequence 186, App
31	637	37.4	364	9	US-10-123-292-186	Sequence 186, App
32	637	37.4	364	9	US-10-123-903-186	Sequence 186, App
33	637	37.4	364	9	US-10-124-819-186	Sequence 186, App
34	637	37.4	364	9	US-10-124-829-186	Sequence 186, App
35	637	37.4	364	9	US-10-140-925-186	Sequence 186, App
36	637	37.4	364	9	US-10-160-498-186	Sequence 186, App
37	637	37.4	364	9	US-10-121-041-186	Sequence 186, App
38	637	37.4	364	9	US-10-121-043-186	Sequence 186, App
39	637	37.4	364	9	US-10-121-047-186	Sequence 186, App
40	637	37.4	364	9	US-10-123-215-186	Sequence 186, App
41	637	37.4	364	9	US-10-123-908-186	Sequence 186, App
42	637	37.4	364	9	US-10-123-908-186	Sequence 186, App
43	637	37.4	364	9	US-10-123-909-186	Sequence 186, App
44	637	37.4	364	9	US-10-123-910-186	Sequence 186, App
45	637	37.4	364	9	US-10-124-813-186	Sequence 186, App

#### ALIGNMENTS

RESULT 1  
US-10-083-853-2  
Sequence 2, Application US/10083853  
Patent No. US20020164709A1  
GENERAL INFORMATION:  
APPLICANT: Affymetrix, Inc  
APPLICANT: Shigeta, Ron T  
APPLICANT: Shigeta, Michael A  
TITLE OF INVENTION: Nucleic Acid Encoding Growth Factor Protein  
FILE REFERENCE: 3385.1  
CURRENT APPLICATION NUMBER: US/10/083, 853  
CURRENT FILING DATE: 2002-02-26  
PRIOR APPLICATION NUMBER: USSN 60/272,663  
PRIOR FILING DATE: 2001-03-01  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: Patent version 3.1  
SEQ ID NO 2  
LENGTH: 317  
TYPE: PRT  
ORGANISM: Homo Sapiens  
US-10-083-853-2

Query Match	100.0%	Score 1704	DB 9	Length 317
Best Local Similarity	100.0%	Pred. No. 4.2e-136		
Matches 317	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	MGDFPMKTPKAMATYAKIDKMDLTKKSFCTAKETTRIVNRQLTWKTFATYSPDKGL	60	
DB	1	MGDFPMKTPKAMATYAKIDKMDLTKKSFCTAKETTRIVNRQLTWKTFATYSPDKGL	60	
QY	61	ISRIYELKQIYKKTKNPKIKVWDMNRFSEGIYAAKKHKKYSSSLATREMOIKTT	120	
DB	61	ISRIYELKQIYKKTKNPKIKVWDMNRFSEGIYAAKKHKKYSSSLATREMOIKTT	120	
QY	121	MRHYLTPVRMAIIKSGNNRDMDEAGNHSQQTITRTKQTPHVLTHRWLLOQSHWTVL	180	
DB	121	MRHYLTPVRMAIIKSGNNRDMDEAGNHSQQTITRTKQTPHVLTHRWLLOQSHWTVL	180	
QY	181	SDISELMHKTDRIVNLIMCNYLLTVDLRLNDAKRYSCTPRYSVNIRELKLAVVFF	240	
DB	181	SDISELMHKTDRIVNLIMCNYLLTVDLRLNDAKRYSCTPRYSVNIRELKLAVVFF	240	

Qy 241 PRCILVORCGNCGGTVMNRSCNCGKTVKCYHEVLQFEPGHIKRRGRKTAALVDIQ 300  
Db 241 PRCILVORCGNCGGTVMNRSCNCGKTVKCYHEVLQFEPGHIKRRGRKTAALVDIQ 300  
Qy 301 LDHHERCDICSSRPFR 317  
Db 301 LDHHERCDICSSRPFR 317

RESULT 2  
US-10-000-256A-153

; Sequence 153, Application US/10000256A  
; Publication No. US20030039983A1  
; GENERAL INFORMATION:  
; APPLICANT: Sun, Yongming  
; APPLICANT: Recipon, Heve  
; APPLICANT: Chen, Sei-Yu  
; APPLICANT: Liu, Chenghua  
; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific Genes and  
; FILE REFERENCE: DEX-0259  
; CURRENT APPLICATION NUMBER: US/10/000,256A  
; PRIOR FILING DATE: 2001-11-01  
; PRIOR APPLICATION NUMBER: 60/244,782  
; PRIOR FILING DATE: 2000-11-01  
; NUMBER OF SEQ ID NOS: 240  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 153  
; LENGTH: 956  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-10-000-256A-153

Query Match 39.9%; Score 680.5; DB 9; Length 956;  
Best Local Similarity 72.9%; Pred. No. 3.4e-49;  
Matches 148; Conservative 5; Mismatches 29; Indels 21; Gaps 5;  
Qy 1 MGKDFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQUTWEKIFATYSPDKGL 60  
Db 671 MGKDFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQUTWEKIFATYSPDKGL 730  
Qy 61 ISRIYNELKQIYKKKTNPPIKKWVKDNRHFSKEGIYAAKHKMKYSSSLAIREMOIKTT 120  
Db 731 ISRIYNELKQIYKKKTNPPIKKWVKDNRHFSKEGIYAAKHKMKYSSSLAIREMOIKTT 790  
Qy 121 MRHYLTPVRMAIITKSGNNRDMDEAGNHSSQOITTRTKNOTPHVLTIRW---IIQOSHVV 177  
Db 791 MRHYLTPVRMAIITKSGNNRDMDEAGNHSSQOITTRTKNOTPHVLTIRW---IIQOSHVV 836  
Qy 178 TV---LSDISELMHKTDRIYNLL 197  
Db 837 SVMRFRLDL-ELRIIPDPAIPL 858

## RESULT 3

US-10-001-835-189  
; Sequence 189, Application US/10001835  
; Patent No. US20020160387A1  
; GENERAL INFORMATION:  
; APPLICANT: Salceda, Susana  
; APPLICANT: Macina, Roberto  
; APPLICANT: Recipon, Heve  
; APPLICANT: Cafferey, Robert  
; APPLICANT: Sun, Yongming  
; APPLICANT: Liu, Chenghua  
; TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and Pro  
; FILE REFERENCE: DEX-0277  
; CURRENT APPLICATION NUMBER: US/10/001,835  
; PRIOR FILING DATE: 2001-11-20  
; PRIOR APPLICATION NUMBER: 60/249,997  
; PRIOR FILING DATE: 2000-11-20  
; NUMBER OF SEQ ID NOS: 228  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 189

; LENGTH: 917  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-10-001-835-189

Query Match 39.5%; Score 673; DB 9; Length 917;  
Best Local Similarity 94.2%; Pred. No. 1.4e-48;  
Matches 131; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 1 MGKDFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQUTWEKIFATYSPDKGL 60  
Db 779 VGKDFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQUTWEKIFATYSPDKGL 838  
Qy 61 ISRIYNELKQIYKKKTNPPIKKWVKDNRHFSKEGIYAAKHKMKYSSSLAIREMOIKTT 120  
Db 839 ISRIYNELKQIYKKKTNPPIKKWVKDNRHFSKEGIYAAKHKMKYSSSLAIREMOIKTT 898  
Qy 121 MRHYLTPVRMAIITKSGNN 139  
Db 899 MRHYLTPVRMAIITKSGNN 917

RESULT 4  
US-10-025-201-3

; Sequence 3, Application US/10025201  
; Publication No. US20030003468A1  
; GENERAL INFORMATION:  
; APPLICANT: Crow, Mary K.  
; TITLE OF INVENTION: MARKERS FOR DISEASE SUSCEPTIBILITY AND TARGETS FOR THERAPY  
; FILE REFERENCE: 5983/2H567  
; CURRENT APPLICATION NUMBER: US/10/025,201  
; PRIOR FILING DATE: 2001-12-19  
; PRIOR APPLICATION NUMBER: 60/256,673  
; PRIOR FILING DATE: 2000-12-19  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 3  
; LENGTH: 1275  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; DATABASE ACCESSION NUMBER: GenBank Accession No. US20030003468A1 U09116  
; DATABASE ENTRY DATE: 1995-02-02  
; RELEVANT RESIDUES: (1)..  
US-10-025-201-3

Query Match 39.5%; Score 672.5; DB 9; Length 1275;  
Best Local Similarity 71.4%; Pred. No. 2.3e-48;  
Matches 147; Conservative 4; Mismatches 28; Indels 27; Gaps 5;

Qy 1 MGKDFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQUTWEKIFATYSPDKGL 60  
Db 990 VGKDFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQUTWEKIFATYSPDKGL 1049  
Qy 61 ISRIYNELKQIYKKKTNPPIKKWVKDNRHFSKEGIYAAKHKMKYSSSLAIREMOIKTT 120  
Db 1050 ISRIYNELKQIYKKKTNPPIKKWVKDNRHFSKEGIYAAKHKMKYSSSLAIREMOIKTT 1109  
Qy 121 MRHYLTPVRMAIITKSGNN--RDMDAGNHSSQOITTRTKNOTPHVLTIRW---IIQOSHVV 174  
Db 1110 MRHYLTPVRMAIITKSGNNRCWRGCGEIG-----TLHCWMDCKLAQPLVK 1152  
Qy 175 HWYTV---LSDISELMHKTDRIYNLL 197  
Db 1153 LKSWRFRLDL-ELRIIPDPAIPL 1177

## RESULT 5

US-10-086-623-6  
; Sequence 6, Application US/10086623  
; Patent No. US20020164710A1  
; GENERAL INFORMATION:  
; APPLICANT: ERIKSSON, ULF

```

: APPLICANT: AASE, Karin
: APPLICANT: Li, Xuri
: APPLICANT: PONTEN, Annica
: APPLICANT: TUTTELA, Marko
: APPLICANT: ALTTALO, Kari
: APPLICANT: OESTMAN, Arne
: APPLICANT: HEIDIN, Carl Henrik
: TITLE OF INVENTION: PLATELET DERIVED GROWTH FACTOR D, DNA CODING THEREFOR AND USES THEREOF
: FILE REFERENCE: 1064/44833C2
: CURRENT APPLICATION NUMBER: US/10/086,623
: PRIOR FILING DATE: 2000-03-04
: PRIOR APPLICATION NUMBER: US 60/107,852
: PRIOR FILING DATE: 1998-11-10
: PRIOR APPLICATION NUMBER: US 60/113,997
: PRIOR FILING DATE: 1998-12-28
: PRIOR APPLICATION NUMBER: US 60/150,604
: PRIOR FILING DATE: 1999-08-26
: PRIOR APPLICATION NUMBER: US 60/157,108
: PRIOR FILING DATE: 1999-10-04
: PRIOR APPLICATION NUMBER: US 60/157,756
: PRIOR FILING DATE: 1999-10-05
: PRIOR APPLICATION NUMBER: US 09/438,046
: PRIOR FILING DATE: 1999-11-10
: PRIOR APPLICATION NUMBER: US 09/691,200
: PRIOR FILING DATE: 2000-10-19
: NUMBER OF SEQ ID NOS: 42
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 6
: LENGTH: 322
: TYPE: prt
: ORGANISM: Homo sapiens
: US-10-086-623-6

```

```

Query Match          37.4%; Score 637; DB 9; Length 322;
Best Local Similarity 65.4%; Pred. No. 4.5e-46;
Matches 125; Conservative 9; Mismatches 11; Indels 46; Gaps 3;

QY      173 QSHWTVVJSDISELMHKT-----DRIVNLMC-----M 200
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      132 ETNNESVTSSISGSYNSPSVTDPTLADLDKKIAEDTVEDLLKYNPESMQEDLENM 191
QY      201 YLT-----VDLDRLNDKARYSCTPRNYSVNIIBELKIANVFFPRCLLV 246
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      192 YLDTPRYKRGSRVHDRKSKVDDLRLNDADKARYSCPRNYSVNIIBELKIANVFFPRCLLV 251
QY      247 QRCGAGCCCGGVNMRSCCTCNSGKTYKKTHEVLOEPFGHKKRGRAKTALVDIOLDHHR 306
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      252 QRCGNCGCCGVNMRSCCTCNSGKTYKKTHEVLOEPFGHKKRGRAKTALVDIOLDHHR 311
QY      307 CDCICSSRPPR 317
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      312 CDCICSSRPPR 322

RESULT 6
US-10-260-539-6
; Sequence 6, Application US/10260539
; Publication No. US20030073637A1
; GENERAL INFORMATION:
; APPLICANT: ERIKSSON, ULF
; APPLICANT: AASE, Karin
; APPLICANT: LI, Kuri
; APPLICANT: PONTEN, Annica
; APPLICANT: TUTELA, Marko
; APPLICANT: ALITALO, Kari
; APPLICANT: OESTMAN, Arne
; APPLICANT: HELDIN, Carl-Henrik
; TITLE OF INVENTION: PLATELET DERIVED GROWTH FACTOR D, DNA CODING THEREFOR AND USES TH
; FILE REFERENCE: 1064/44833C2
; CURRENT APPLICATION NUMBER: US/10/260.539
; CURRENT FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: US/10/086.623
; PRIOR FILING DATE: 2000-03-04

```

```

1 PRIOR APPLICATION NUMBER: US 60/107,882
2 PRIOR FILING DATE: 1998-11-10
3 PRIOR APPLICATION NUMBER: US 60/113,9937
4 PRIOR FILING DATE: 1998-12-28
5 PRIOR APPLICATION NUMBER: US 60/150,604
6 PRIOR FILING DATE: 1999-08-26
7 PRIOR APPLICATION NUMBER: US 60/157,108
8 PRIOR FILING DATE: 1999-10-04
9 PRIOR APPLICATION NUMBER: US 60/157,756
10 PRIOR FILING DATE: 1999-10-05
11 PRIOR APPLICATION NUMBER: US 07/438,046
12 PRIOR FILING DATE: 1999-11-10
13 PRIOR APPLICATION NUMBER: US 09/691,200
14 PRIOR FILING DATE: 2000-10-19
15 NUMBER OF SEQ ID NOS: 42
16 SOFTWARE: PatentIn version 3.1
17 SEQ ID NO 6
18 LENGTH: 322
19 TYPE: PRT
20 ORGANISM: Homo sapiens
21 US-10-260-839-6

```

Query Match	37.4%	Score 637;	DB 9;	Length 322;
Best Local Similarity	65.4%;	Pred. No. 4.5e-46;		
Matches 125; Conservative	9;	Mismatches 11;	Indels 46;	Gaps 3;

[illegible]

RESULT 7  
US-10-028-072-186  
Sequence 186, Application US/10028072  
Publication No. US2003000431A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Beresini, Maureen  
APPLICANT: DeForge, Laura  
APPLICANT: Desnoyers, Luc  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerlitsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Sherwood, Steven  
APPLICANT: Smith, Victoria  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tunas, Daniel  
APPLICANT: Watanabe, Colin K  
APPLICANT: Wood, William  
APPLICANT: Zhang  
TITLE OR INVENTION:  
FILE REFERENCE:  
CURRENT APPLICATION NUMBER: US/10/028, 072  
CURRENT FILING DATE: 2001-12-19  
PRIOR APPLICATION NUMBER: 60/049911  
PRIOR FILING DATE: 1997-06-18  
PRIOR APPLICATION NUMBER: 60/056974  
PRIOR FILING DATE: 1997-08-26





```

; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 19/98-06-11
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090538
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07

```

```

Query Match      37.4%; Score 637; DB 9; Length 364;
Best Local Similarity 65.4%; Pred. No. 5.2e-46;
Matches 125; Conservative 9; Mismatches 11; Indels 46; Gaps 3;

```

```

QY 173 QSHWTVLSDISLHMKT-----DRIVNLMC-----M 200
DB 174 ETWESVTSISGVSNPSVTPTLADLADPKIAEFVTDLLKYFNPESNQEDLENN 233
QY 201 YLLT-----VDLRLNDAKRYSCPTPNYSVNIREELKLANVFFPRCLV 246
DB 234 YLDTPRYGRSYHDKSKVDLRLNDAKRYSCPTPNYSVNIREELKLANVFFPRCLV 233
QY 247 QRCGNGCGGTVMWRSCTNSGKTVKYEVLQFEPGHIKRRGRATMALVDIQLDHER 306
DB 294 QRCGNGCGGTVMWRSCTNSGKTVKYEVLQFEPGHIKRRGRATMALVDIQLDHER 353
QY 307 CDCICSSRPPR 317
DB 354 CDCICSSRPPR 364

```

## RESULT 8

```

US-10-121-049-186
; Sequence 186, Application US/10121049
; Publication No. US2003002239A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerlitsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumaas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C17

```

```

; CURRENT APPLICATION NUMBER: US/10/121, 049
; CURRENT FILING DATE: 2002-04-12
; Prior Application removed - See file wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 186
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-121-049-186

```

```

Query Match      37.4%; Score 637; DB 9; Length 364;
Best Local Similarity 65.4%; Pred. No. 5.2e-46;
Matches 125; Conservative 9; Mismatches 11; Indels 46; Gaps 3;

```

```

QY 173 QSHWTVLSDISLHMKT-----DRIVNLMC-----M 200
DB 174 ETWESVTSISGVSNPSVTPTLADLADPKIAEFVTDLLKYFNPESNQEDLENN 233
QY 201 YLLT-----VDLRLNDAKRYSCPTPNYSVNIREELKLANVFFPRCLV 246
DB 234 YLDTPRYGRSYHDKSKVDLRLNDAKRYSCPTPNYSVNIREELKLANVFFPRCLV 233
QY 247 QRCGNGCGGTVMWRSCTNSGKTVKYEVLQFEPGHIKRRGRATMALVDIQLDHER 306
DB 294 QRCGNGCGGTVMWRSCTNSGKTVKYEVLQFEPGHIKRRGRATMALVDIQLDHER 353
QY 307 CDCICSSRPPR 317
DB 354 CDCICSSRPPR 364

```

## RESULT 9

```

US-10-123-904-186
; Sequence 186, Application US/10123904
; Publication No. US20030022328A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerlitsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumaas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C54
; CURRENT APPLICATION NUMBER: US/10/123, 904
; CURRENT FILING DATE: 2002-04-16
; Prior Application removed - See file wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 186
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-123-904-186

```

```

Query Match      37.4%; Score 637; DB 9; Length 364;
Best Local Similarity 65.4%; Pred. No. 5.2e-46;
Matches 125; Conservative 9; Mismatches 11; Indels 46; Gaps 3;

```

```

QY 173 QSHWTVLSDISLHMKT-----DRIVNLMC-----M 200
DB 174 ETWESVTSISGVSNPSVTPTLADLADPKIAEFVTDLLKYFNPESNQEDLENN 233

```

```
QY 201 YLTT-----VDLRLNDADAKRYSCTPRNVSVNIREELKLANVFFPRCLLV 246
D 234 YLDTPRYRGRSYHDRKSKVDLRLNDADAKRYSCTPRNVSVNIREELKLANVFFPRCLLV 293
QY 247 QRCGNGCGCTVWNRSTCNSGKTVKKYHEVLOFEFGHKKRGRAKTMALVDIOLDHHR 306
D 294 QRCGNGCGCTVWNRSTCNSGKTVKKYHEVLOFEFGHKKRGRAKTMALVDIOLDHHR 353
QY 307 CDCICSSRPPR 317
D 354 CDCICSSRPPR 364

RESULT 10
US-10-140-470-186
; Sequence 186, Application US/10140470
; Publication No. US20030022331A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerlitsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Matanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Zhan, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P330R1C160
; CURRENT APPLICATION NUMBER: US/10/140,470
; PRIOR FILING DATE: 2002-05-06
; PRIOR APPLICATION REMOVED - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 186
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-470-186

Query Match 37.4%; Score 637; DB 9; Length 364;
Best Local Similarity 65.4%; Pred. No. 5,2e-46;
Matches 125; Conservative 9; Mismatches 11; Indels 46; Gaps 3;

QY 173 QSHWVTYLSIDISELMHKT-----DRIVNLMC-----M 200
D 174 ETNWSVTSSISGVSNPSVTDPTLLADALDKIAEFTVEDLKYFNPSWQEDLENN 233
QY 201 YLTT-----VDLRLNDADAKRYSCTPRNVSVNIREELKLANVFFPRCLLV 246
D 234 YLDTPRYRGRSYHDRKSKVDLRLNDADAKRYSCTPRNVSVNIREELKLANVFFPRCLLV 293
QY 247 QRCGNGCGCTVWNRSTCNSGKTVKKYHEVLOFEFGHKKRGRAKTMALVDIOLDHHR 306
D 294 QRCGNGCGCTVWNRSTCNSGKTVKKYHEVLOFEFGHKKRGRAKTMALVDIOLDHHR 353
QY 307 CDCICSSRPPR 317
D 354 CDCICSSRPPR 364

RESULT 11
US-10-175-746-186
; Sequence 186, Application US/10175746
```

```
Publication No. US20030027270A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerlitsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Matanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Zhan, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P330R1C153
; CURRENT APPLICATION NUMBER: US/10/175,746
; PRIOR FILING DATE: 2002-06-19
; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 186
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-746-186

Query Match 37.4%; Score 637; DB 9; Length 364;
Best Local Similarity 65.4%; Pred. No. 5,2e-46;
Matches 125; Conservative 9; Mismatches 11; Indels 46; Gaps 3;

QY 173 QSHWVTYLSIDISELMHKT-----DRIVNLMC-----M 200
D 174 ETNWSVTSSISGVSNPSVTDPTLLADALDKIAEFTVEDLKYFNPSWQEDLENN 233
QY 201 YLTT-----VDLRLNDADAKRYSCTPRNVSVNIREELKLANVFFPRCLLV 246
D 234 YLDTPRYRGRSYHDRKSKVDLRLNDADAKRYSCTPRNVSVNIREELKLANVFFPRCLLV 293
QY 247 QRCGNGCGCTVWNRSTCNSGKTVKKYHEVLOFEFGHKKRGRAKTMALVDIOLDHHR 306
D 294 QRCGNGCGCTVWNRSTCNSGKTVKKYHEVLOFEFGHKKRGRAKTMALVDIOLDHHR 353
QY 307 CDCICSSRPPR 317
D 354 CDCICSSRPPR 364

RESULT 12
US-10-176-918-186
; Sequence 186, Application US/10176918
; Publication No. US20030027275A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerlitsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Matanabe, Colin K
```

```

; APPLICANT: Wood,William
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C382
; CURRENT APPLICATION NUMBER: US/10/176,918
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 186
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-918-186

```

```

Query Match 37.4%; Score 637; DB 9; Length 364;
Best Local Similarity 65.4%; Pred. No. 5.2e-46;
Matches 125; Conservative 9; Mismatches 11; Indels 46; Gaps 3;

```

```

QY 173 QSHWTVLSDISELMHKT-----DRIVNLMC-----M 200
DB 174 ETWVESVTSISGVSNSSVTPDTLIADALDKIAEFVTEVDLKYFNPESWQEDLENN 233
QY 201 YLLT-----VDLRINDAKRYGCTPNYSVNIREEKLNAVVFPPCLIV 246
DB 234 YLDTPRYGRSYHDKRSKVDLRINDAKRYGCTPNYSVNIREEKLNAVVFPPCLIV 293
QY 247 ORCGNCGGCTVWNRSGCTNSGKTVKYEVLQFEPGHIKRRGRATMALVDIQDHHHR 306
DB 294 ORCGNCGGCTVWNRSGCTNSGKTVKYEVLQFEPGHIKRRGRATMALVDIQDHHHR 353
QY 307 CDCICSSRPPR 317
DB 354 CDCICSSRPPR 364

```

```

RESULT 13
US-10-176-921-186
; Sequence 186, Application US/10176921
; Publication No. US20030027276A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C288
; CURRENT APPLICATION NUMBER: US/10/176,921
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 186
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-921-186

```

```

Query Match 37.4%; Score 637; DB 9; Length 364;
Best Local Similarity 65.4%; Pred. No. 5.2e-46;

```

```

Matches 125; Conservative 9; Mismatches 11; Indels 46; Gaps 3;
QY 173 QSHWTVLSDISELMHKT-----DRIVNLMC-----M 200
DB 174 ETWVESVTSISGVSNSSVTPDTLIADALDKIAEFVTEVDLKYFNPESWQEDLENN 233
QY 201 YLLT-----VDLRINDAKRYGCTPNYSVNIREEKLNAVVFPPCLIV 246
DB 234 YLDTPRYGRSYHDKRSKVDLRINDAKRYGCTPNYSVNIREEKLNAVVFPPCLIV 293
QY 247 ORCGNCGGCTVWNRSGCTNSGKTVKYEVLQFEPGHIKRRGRATMALVDIQDHHHR 306
DB 294 ORCGNCGGCTVWNRSGCTNSGKTVKYEVLQFEPGHIKRRGRATMALVDIQDHHHR 353
QY 307 CDCICSSRPPR 317
DB 354 CDCICSSRPPR 364

```

```

RESULT 14
US-10-137-865-186
; Sequence 186, Application US/10137865
; Publication No. US20030032155A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C154
; CURRENT APPLICATION NUMBER: US/10/137,865
; CURRENT FILING DATE: 2002-05-03
; Prior Application removed - See File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 186
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-137-865-186

```

```

Query Match 37.4%; Score 637; DB 9; Length 364;
Best Local Similarity 65.4%; Pred. No. 5.2e-46;
Matches 125; Conservative 9; Mismatches 11; Indels 46; Gaps 3;

```

```

QY 173 QSHWTVLSDISELMHKT-----DRIVNLMC-----M 200
DB 174 ETWVESVTSISGVSNSSVTPDTLIADALDKIAEFVTEVDLKYFNPESWQEDLENN 233
QY 201 YLLT-----VDLRINDAKRYGCTPNYSVNIREEKLNAVVFPPCLIV 246
DB 234 YLDTPRYGRSYHDKRSKVDLRINDAKRYGCTPNYSVNIREEKLNAVVFPPCLIV 293
QY 247 ORCGNCGGCTVWNRSGCTNSGKTVKYEVLQFEPGHIKRRGRATMALVDIQDHHHR 306
DB 294 ORCGNCGGCTVWNRSGCTNSGKTVKYEVLQFEPGHIKRRGRATMALVDIQDHHHR 353
QY 307 CDCICSSRPPR 317
DB 354 CDCICSSRPPR 364

```

RESULT 15

US-10-140-474-186

; Sequence 186, Application US/10140474  
; Publication No. US20030032156A1  
; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: Deforge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gueney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Thomas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; FILE REFERENCE: P330R1C162

; CURRENT APPLICATION NUMBER: US/10/140,474

; CURRENT FILING DATE: 2002-05-06

; Prior Application removed - See Palm or File Wrapper

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 186

; LENGTH: 364

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-140-474-186

Query March 37.4%; Score 637; DB 9; Length 364;

Best Local Similarity 65.4%; Pred. No. 5.2e-46;

Matches 125; Conservative 9; Mismatches 11; Indels 46; Gaps 3;

QY 173 QSHWTVVLSDISLHMKT-----DRIVLLMC-----M 200

Db 174 ETWSEYTSISGVSYNSPVTDPTLADALDKKIAEFDVIEDLLKYFNPEMQEDLENM 233

QY 201 YLLT-----VDRLNDKAKRYSCPTPNYSVNIIEELKLANVFFPRLCLV 246

Db 234 YLDTPRYGRGRSYNDKSKVDLDRLNDKAKRYSCPTPNYSVNIIEELKLANVFFPRLCLV 293

QY 247 QRCGNGCGCGTVMNRSCCTCNSGKTVMKKYHEVLOFEPGHIKRGRAKTMALVDIOLDHHR 306

Db 294 QRCGNGCGCGTVMNRSCCTCNSGKTVMKKYHEVLOFEPGHIKRGRAKTMALVDIOLDHHR 353

QY 307 CDCICSSRPFR 317

Db 354 CDCICSSRPFR 364

Search completed: June 5, 2003, 12:52:58  
Job time: 48 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2003, 12:32:48 ; Search time 23 Seconds

(without alignments)  
571.652 Million cell updates/sec

Title: US-10-083-853B-1

Sequence: 1 MGDFPMSKTPKMATKAKID.....DIQLDHERDCCISRPFR 317

Scoring table: ELOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892, seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing First 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	658.5	38.6	1259	1	LINI_HUMAN
2	443.5	26.0	1260	1	LINI_NYCCO
3	412.5	24.2	1300	1	POL2_MOUSE
4	125.5	7.4	148	1	VEGH_ORPNT
5	116.5	6.8	326	1	VEGD_RAT
6	114.5	6.7	334	1	VEGD_HUMAN
7	106.5	6.2	358	1	VEGD_MOUSE
8	101	5.9	1576	1	YLK3_CABEL
9	96.5	5.7	419	1	VEGC_HUMAN
10	95.5	5.6	135	1	VEGB_RAT
11	95.5	5.6	207	1	VEGB_MOUSE
12	94	5.5	213	1	PDGA_RABIT
13	92.5	5.4	4092	1	DYHC_YEAST
14	92	5.4	133	1	VEGH_ORPNT
15	92	5.4	415	1	VEGC_MOUSE
16	91	5.3	803	1	GYRE_BUCAL
17	89.5	5.3	437	1	CSN2_SCHPO
18	89.5	5.3	704	1	DBF4_YEAST
19	89	5.2	204	1	PDGA_RAT
20	89	5.2	211	1	PDGA_MOUSE
21	89	5.2	211	1	ULP1_YEAST
22	88.5	5.2	621	1	GOG4_HUMAN
23	88.5	5.2	1230	1	VEGA_HORSE
24	88	5.2	190	1	VEGA_HORSE
25	87.5	5.1	298	1	FORB_METJA
26	87	5.1	1202	1	DPOW_ASCIM
27	87	5.1	1226	1	KFA4_XENLA
28	87	5.1	2349	1	TPR_HUMAN
29	87	5.1	4349	1	DYHC_FUSO
30	86.5	5.1	207	1	VEGB_BOVIN
31	86.5	5.1	1312	1	RASO_YEAST
32	86.5	5.1	3391	1	POLG_DEN26
33	86	5.0	1531	1	YQ38_CABEL

34	85.5	5.0	738	1	YKF4_YEAST	P35732 saccharomyc
35	85	5.0	740	1	DDX1_HUMAN	O92499 homo sapien
36	85	5.0	1038	1	CIN8_YEAST	P27895 saccharomyc
37	84.5	5.0	338	1	Y941_METJA	O57711 methanococc
38	84.5	5.0	578	1	OAR2_LYMTST	O01670 lymnaea sta
39	84.5	5.0	596	1	PRIM_CLOAB	P33655 clostridium
40	84	4.9	394	1	KIM4_HUMAN	O76011 homo sapien
41	84	4.9	1327	1	YKMA_CABEL	P34544 caenorhabdi
42	83.5	4.9	207	1	VEGB_HUMAN	P49765 homo sapien
43	83.5	4.9	486	1	UN53_ROTID	P35423 human rotav
44	83.5	4.9	789	1	A12M_YEAST	P03876 saccharomyc
45	83	4.9	226	1	PDGA_XENLA	P13698 xenopus lae

## ALIGNMENTS

RESULT 1	LINE1_HUMAN	STANDARD;	PRT; 1259 AA.
AC	P08547;		
DT	01-AUG-1988 (Rel. 08, Created)		
DT	01-AUG-1988 (Rel. 08, Last sequence update)		
DT	01-AUG-1988 (Rel. 08, Last annotation update)		
DE	LINE-1 reverse transcriptase homolog.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE.		
RX	MEDLINE=86230917; PubMed=2423883;		
RA	Hattori M., Kuhara S., Takenaka O., Sakaki Y.;		
RT	"A family of repetitive DNA sequences in primates may be derived		
RT	from a sequence encoding a reverse transcriptase-related protein."		
RL	Nature 321:625-628 (1986).		
CC	-1- MISCELLANEOUS: THIS SEQUENCE WAS CONSTRUCTED FROM AN ALIGNMENT OF		
CC	PUBLISHED AND UNPUBLISHED SEQUENCES, DETERMINED IN VARIOUS		
CC	LABORATORIES, BELONGING TO THE LINE-1 FAMILY.		
DR	PIR; A25313; GNMHUL.		
DR	InterPro; IPR005135; Exo_endo_phos.		
DR	InterPro; IPR000300; IPFC.		
DR	InterPro; IPR000477; RVTse.		
DR	Pfam; PF000078; rvt; 1.		
DR	Pfam; PF03372; Exo_endo_phos; 1.		
DR	SMART; SM00128; IPFC; 1.		
KW	RNA-directed DNA polymerase.		
SO	SEQUENCE 1259 AA; 147112 MW; F3BC18A3803919F5 CRC64;		

Query Match	38.6%;	Score 658.5;	DR 1;	Length 1259;
Best Local Similarity	69.9%;	Pred. No. 3,4e-45;		
Matches 144;	Conservative 6;	Mismatches 29;	Indels 27;	Gaps 5;

QY	1	MGDFPMSKTPKMATKAKIDKMDLKLKSFCTAKETTTIVNROLTWEKIFATYSPDKGL	60
DB	989	MGDFPMSKTPKMATKAKIDKMDLKLKSFCTAKETTTIVNROLTWEKIFATYSPDKGL	1048
QY	61	ISRIYVELKQIYKKTNDPKKRVKQWNRHFSKEGIYAAKKNKXSSSLAIREMOIKTT	120
DB	1049	ISRIYVELKQIYKKTNDPKKRVKQWNRHFSKEGIYAAKKNKXSSSLAIREMOIKTT	1108
QY	121	MRVHLTPVMAIIKKSGNN---RDMDEAGNHSSQGITTTTKNQTPHYLTHRW---ILQGS	174
DB	1109	MRVHLTPVMAIIKKSGNNRMRGCGEIG-----TLHICWMDCKLVOP	1151
QY	175	HWVTV---LSDISELMHKTDIRIVNLL	197
DB	1152	LKMSVWRFLRDL-ELEIIPDPAIPII	1176

RESULT 2	LINE1_NYCCO	STANDARD;	PRT; 1260 AA.
----------	-------------	-----------	---------------

```

AC P08548;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-AUG-1988 (Rel. 08, Last annotation update)
DE LINE-1 reverse transcriptase homolog.
OS Nucleobase coucang (Slow Loris).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Strepsirhini; Loridae; Nycticebus.
OX NCBI_TaxID=9470;
RN [1]
RP SEQUENCE.
RX MEDLINE=86230917; PubMed=2423883;
RA Hatori M., Kuhara S., Takenaka O., Sakaki Y.;
RT "LI family of repetitive DNA sequences in primates may be derived
RT from a sequence encoding a reverse transcriptase-related protein."
RL Nature 321:625-628(1986)
CC -I- MISCELLANEOUS: THIS SEQUENCE WAS CONSTRUCTED FROM AN ALIGNMENT OF
CC SIX SEQUENCES, DETERMINED BY THESE AUTHORS BUT NOT SHOWN.
CC BELONGING TO THE LINE-1 FAMILY.
CC PIR; B25313; GNLRL1.
DR HSSP; P27695; 1HD7.
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR00477; RVTse.
DR Pfam; PF00078; rvt; 1.
DR Pfam; PF03372; Exo_endo_phos; 1.
KW RNA-directed DNA polymerase.
SQ SEQUENCE 1260 AA; 147042 MW; 7A6803DF471F7253 CRC64;

Query Match 26.0%; Score 443.5; DB 1; Length 1260;
Best Local Similarity 50.3%; Pred. No. 6.5e-28;
Matches 95; Conservative 25; Mismatches 52; Indels 17; Gaps 2;

QY 1 MGKQFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQLTEWEKIFATYSPDKGL 60
DB 989 LGEFMRRTPAIEVSKIKHWDLIKLSFCTAKIYKASRQSEWKEITAGTISDKGL 1048
QY 61 ISRIYNELKQIKKTKNPIKWKVDMNHRHFSKEGIYAKKMKKYSLSLAIREMOIKTT 120
DB 1049 ITRIRHRELDKHKIKTRDPISGMARDLKRNFSEKEDRHITIKMKSSSLIRERQIKTT 1108
QY 121 MRHYLTPRMAIIKKSNNRMDENGNHNSOOTITRTNQPHYLTHRW---IIQSHWV 177
DB 1109 LRYHLTPRVAHITKSPNQRCWRCGGKGT-----LLHCWMECPILRSPWK 1154
QY 178 TVLSDISEL 186
DB 1155 DVWRILRDL 1163

RESULT 3
POL2_MOUSE STANDARD; PRT; 1300 AA.
ID POL2_MOUSE
AC P11369;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Retrovirus-related POL polypeptide [contains: Reverse transcriptase
DE (EC 2.7.7.49); Endonuclease].
GN POL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87064284; PubMed=3023821;
RA Leob D.D., Padgett R.W., Hardies S.C., Shehee W.R., Comer M.B.,
RA Edgell M.H., Hutchison C.A. III;
RT "The sequence of a large Lmd element reveals a tandemly repeated 5'
RT end and several features found in retrotransposons."
RL Mol. Cell. Biol. 6:168-182(1986).
CC -I- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA} (N).

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M13002; AAA66024.1; ALT_INIT.
DR PIR; B24906; GNMSSL.
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR00477; RVTse.
DR Pfam; PF00078; rvt; 1.
DR Pfam; PF03372; Exo_endo_phos; 1.
KW Hydroxylase; Transferase; RNA-directed DNA polymerase; Nuclease;
KW Endonuclease; Polypeptide.
SQ SEQUENCE 1300 AA; 151829 MW; 9EE8108493B89635 CRC64;

Query Match 24.2%; Score 412.5; DB 1; Length 1300;
Best Local Similarity 47.0%; Pred. No. 2.1e-25;
Matches 85; Conservative 30; Mismatches 49; Indels 17; Gaps 2;

QY 2 GKQFMSKTPKAMATKAKIDKMDLIKLSFCTAKETTRVNRQLTEWEKIFATYSPDKGL 61
DB 1017 GEKFLNTAMACAVRSRIDKMDIKLQSFCKAKDTVTKRPPDWERIFTPKSDGLI 1076
QY 62 SRIYNELKQIKKTKNPIKWKVDMNHRHFSKEGIYAKKMKKYSLSLAIREMOIKTTM 121
DB 1077 SNIKELKQVDFRKNPIKWKSELSKEFPEYRMAEHLKCCSSLSLIREMOIKTTL 1136
QY 122 RYHLTPRMAIIKKSNNRMDENGNHNSOOTITRTNQPHYLTHRW---IIQSHWV 178
DB 1137 RHYLTPRMAKIKNSGSRRCMGGERGT-----LLHCWMECPILVPLMK 1182
QY 179 V 179
DB 1183 V 1183

RESULT 4
VEGH_ORFNT STANDARD; PRT; 148 AA.
ID VEGH_ORFNT
AC P52585;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Vascular endothelial growth factor homolog precursor.
GN A2R.
OS Orf virus (strain NZ7) (OV NZ-7).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Parapoxvirus.
OX NCBI_TaxID=73495;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94076465; PubMed=8254780;
RA Lytle D.J., Fraser K.M., Fleming S.B., Mercer A.A., Robinson A.J.;
RA "Homologs of vascular endothelial growth factor are encoded by the
RT poxvirus orf virus."
RL J. Virol. 68:84-92(1994).
CC -I- FUNCTION: INDUCES ENDOTHELIAL PROLIFERATION.
CC -I- SUBUNIT: HOMODIMER; DISULFIDE-LINKED (BY SIMILARITY).
CC -I- SIMILARITY: BELONGS TO THE PDGF/VEGF FAMILY OF GROWTH FACTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S67522; AAB29223.1; -.

```

```

DR HSSP; P15692; 2VP.
DR InterPro; IPR000072; PD_growth_factor.
DR Pfam; PF00341; PDGF; 1.
DR ProDom; PD001629; PD_growth_factor; 1.
DR SMART; SM00141; PDGF; 1.
DR PROSITE; PS00249; PDGF_1; FALSE_NEG.
DR PROSITE; PS0278; PDGF_2; 1.
DR Mitogen; Growth factor; Glycoprotein; Signal.
KW SIGNAL
FT SIGNAL 1 25
FT CHAIN 26 148
FT DISULFID 46 88
FT DISULFID 77 130
FT DISULFID 81 132
FT DISULFID 71 132
FT DISULFID 80 80
FT CARBOHYD 95 95
SQ SEQUENCE 148 AA; 16078 MW; F0E13BA104CC73F8 CRC64;

Query Match 7.4%; Score 125.5; DB 1; Length 148;
Best Local Similarity 25.5%; Pred. No. 0.0019;
Matches 41; Conservative 18; Mismatches 43; Indels 59; Gaps 6;

QY 193 IVNLNMCVLLTVLDRLNDLND-----AKRYSCPTPNYSVNIREEU-KLANVFFPP 241
DB 10 VVALLICMVLNPECVGSQSDSPSTNDWMRTLBKSGCKPDTVVYLGEEYESTNLQVNP 69
QY 242 RCLLVORCGNCG-----CCTVNRSCST-----CNSGKTVKRYHEVLQPEFG 283
DB 70 RCVTVNRKCGCCGCGDQICATVETNRTVTSVGVSSSGTNSGVSTN----- 118
QY 284 HIKRGRKATMALVDIQLDHERDCI-----CGRPR 317
DB 119 -----LQISVTEHTKDCICGRITTTPTTRER 147

RESULT 5
VEGD RAT STANDARD; PRT; 326 AA.
AC 035251;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Vascular endothelial growth factor D precursor (VEGF-D) (c-fos induced growth factor) (FIGF).
GN FIGF OR VEGFD.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RA Yamada Y., Hirata Y., Nezu J., Shimane M.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Growth factor active in angiogenesis, lymphangiogenesis and endothelial cell growth, stimulating their proliferation and migration and also has effects on the permeability of blood vessels. May function in the formation of the venous and lymphatic vascular systems during embryogenesis, and also in the maintenance of differentiated lymphatic endothelium in adults. Binds and activates VEGFR-3 (Flt4) receptor (By similarity).
CC -1- SUBUNIT: Homodimer; non-covalent and antiparallel (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -1- PTM: Undergoes a complex proteolytic maturation which generates a variety of processed secreted forms with increased activity toward VEGFR-3 and VEGFR-2. VEGF-D first form an antiparallel homodimer linked by disulfide bonds before secretion. The fully processed VEGF-D is composed mostly of two VEGF homology domains (VHDS) bound by non-covalent interactions (By similarity).
CC -1- SIMILARITY: BELONGS TO THE PDGF/VEGF FAMILY OF GROWTH FACTORS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF014827; AAB6557.1; -.
DR HSSP; P15692; 1VPP.
DR InterPro; IPR004153; CXKCX_repeat.
DR InterPro; IPR000072; PD_growth_factor.
DR Pfam; PF00341; PDGF; 1.
DR ProDom; PD001629; PD_growth_factor; 1.
DR SMART; SM00141; PDGF; 1.
DR PROSITE; PS00249; PDGF_1; 1.
DR PROSITE; PS0278; PDGF_2; 1.
DR Mitogen; Growth factor; Glycoprotein; Signal; Repeat;
KW Cleavage on pair of basic residues; Multigene family.
FT SIGNAL 1 21
FT PROPEP 22 93
FT CHAIN 94 210
FT PROPEP 211 326
FT DOMAIN 227 317
FT REPEAT 227 242
FT REPEAT 263 278
FT REPEAT 282 298
FT DISULFID 306 317
FT DISULFID 116 158
FT DISULFID 147 194
FT DISULFID 151 196
FT DISULFID 141 141
FT DISULFID 150 150
FT CARBOHYD 160 160
FT CARBOHYD 190 190
FT CARBOHYD 292 292
SQ SEQUENCE 326 AA; 37112 MW; 1261AFA373596C00 CRC64;

Query Match 6.8%; Score 116.5; DB 1; Length 326;
Best Local Similarity 24.9%; Pred. No. 0.025;
Matches 45; Conservative 25; Mismatches 70; Indels 41; Gaps 8;

QY 150 SGOITRTKNGPHTVLTMMILQOSH-----WVTYLS-----DISELMKHTDRIYVL 196
DB 38 SRSVLRSSQIRASTLEELQVASEDWKMKRCKLKSILANDSRSTSRSTFA-- 95
QY 197 LMCVLLTVLDRLNDARYSCTPNYSVNIREEU-KLANVFFPRLVORCGNCGC 255
DB 96 -ATFYDTEFLKVIDEWMQRTCCSPRETCVAVSELGTTNFFKPCVNVFRCGCGC-- 151
QY 256 GTVNRSCSTNSGCT--VKYIHEV--IQPEGHIKRGKRAKTMALVDIQLDHERDCI 309
DB 152 ---NEBSVNCMTSTSYISKLFELISVPLTSVP-----ELVPEVKIANHNGCKC 196
QY 310 I 310
DB 197 L 197

RESULT 6
VEGD HUMAN STANDARD; PRT; 354 AA.
AC 043915;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Vascular endothelial growth factor D precursor (VEGF-D) (c-fos induced growth factor) (FIGF).
GN FIGF OR VEGFD.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

```

OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=97349118; PubMed=9205122;  
RA Yamada Y., Nezu J.-I., Shimane M., Hiyata Y.;  
RT "Molecular cloning of a novel vascular endothelial growth factor,  
VEGF-D";  
RL Genomics 47:207-216(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=98140120; PubMed=9479493;  
RA Rocchigiani M., Lestingi M., Luddi A., Orlandini M., Franco B.,  
RA Rossi E., Ballabio A., Zuffardi O., Oliviero S.;  
RT "Human Flt4: cloning, gene structure, and mapping to chromosome Xp22.1  
between the PIGA and the GPR genes";  
RL Genomics 47:207-216(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98118549; PubMed=9435229;  
RA Achen M.G., Jelicich M., Kuk E., Maekinen T., Vitali A., Wilks A.F.,  
RA Alitalo K., Stacker S.A., Simpson R.O., Moritz R., Kaplanen T.,  
RA Alitalo K., Achen M.G.;  
RT "Biochemistry of vascular endothelial growth factor-D involves  
proteolytic processing which generates non-covalent homodimers";  
RL J. Biol. Chem. 274:32127-32136(1999).  
CC -1- FUNCTION: Growth factor active in angiogenesis, lymphangiogenesis  
and endothelial cell growth, stimulating their proliferation and  
migration and also has effects on the permeability of blood  
vessels. May function in the formation of the venous and lymphatic  
vascular systems during embryogenesis, and also in the maintenance  
of differentiated lymphatic endothelium in adults. Binds and  
activates VEGFR-2 (Flk1) and VEGFR-3 (Flt4) receptors.  
CC -1- SUBUNIT: Homodimer; non-covalent and antiparallel.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- TISSUE SPECIFICITY: Highly expressed in lung, heart, small  
intestine and fetal lung, and at lower levels in skeletal muscle,  
colon, and pancreas.  
CC -1- PM: Undergoes a complex proteolytic maturation which generates a  
variety of processed secreted forms with increased activity toward  
VEGFR-3 and VEGFR-2. VEGF-D first form an antiparallel homodimer  
linked by disulfide bonds before secretion. The fully processed  
VEGF-D is composed mostly of two VEGF homology domains (VHDs)  
bound by non-covalent interactions.  
CC -1- SIMILARITY: BELONGS TO THE PDGF/VEGF FAMILY OF GROWTH FACTORS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC -----  
DR EMBL; D89630; BA024364.1; -;  
DR EMBL; Y12863; CAA73370.1; -;  
DR EMBL; Y12864; CAA73371.1; -;  
DR EMBL; Y12865; CAA73371.1; JOINED.  
DR EMBL; Y12866; CAA73371.1; JOINED.  
DR EMBL; Y12867; CAA73371.1; JOINED.  
DR EMBL; Y12868; CAA73371.1; JOINED.  
DR EMBL; Y12869; CAA73371.1; JOINED.  
DR EMBL; Y12870; CAA73371.1; JOINED.  
DR EMBL; AJ000185; CAA03942.1; -;

DR	HSSP; P15692; 1VP.	
DR	Genew; HGNC:3708; FIGF.	
DR	MIM; 300091; -	
DR	InterPro; IPR004153; CKXC_repeat.	
DR	InterPro; IPR000072; PD_growth_factor.	
DR	Pfam; PF00341; PDGF; 1.	
DR	Pfam; PF03128; CKXC; 3.	
DR	ProDom; PD01629; PD_growth_factor; 1.	
DR	SMART; SM00141; PDGF; 1.	
DR	PROSITE; PS00249; PDGF_1; 1.	
DR	PROSITE; PS0278; PDGF_2; 1.	
KM	Mitogen; Growth_factor_Glycoprotein; Signal; Repeat; Cleavage on pair of basic residues, Multigene family.	
KM	Signal	1 21
FT	PROPEP	22 88
FT	CHAIN	89 205
FT	PROPEP	206 354
FT	DOMAIN	222 318
FT	REPEAT	222 237
FT	REPEAT	258 273
FT	REPEAT	277 293
FT	REPEAT	301 318
FT	DISULFID	112 153
FT	DISULFID	141 189
FT	DISULFID	142 189
FT	DISULFID	146 191
FT	DISULFID	136 136
FT	DISULFID	145 145
FT	CARBOHD	155 155
FT	CARBOHD	185 185
FT	CARBOHD	287 287
QO	SEQUENCE	354 AA; 40444 MW; 2048D76D9735173E CRC64;

Query Match	Similarity	6.7%: Score 114.5; DB 1; Length 354;
Best Local	Similarity 24.5%: Pred. No. 0.04; Matches 47; Conservative 27; Mismatches 71; Indels 47; Gaps 9	
Qy	145 AGNNH-----SQQITRTKQTPHYLTHRMILQOSH-----WTVL-----SDISE 185	
Db	22 SSNEHGPYKSSOSTLERSEQIQAASLEELLRTTHSEDMKLMRCRLKLSFTSMDSRS 81	
Qy	186 LMAKTDRIVILMCMYLLTYDLDRLNDACKYSCTPPNYSVNIREEI-KLANVFFPPCL 244	
Db	82 ASHSTRPAATFYDIETLKV---IDEMQRQCSPPRETCVEVASLCKSTNTPFKPCV 137	
Qy	245 LVQRCGCGCGCGTVMSSCTNSGKT--VKYHEV---LQPEPHIKRGRAKMALVD 298	
Db	138 NVFPCGGCGC-----NEESLICMNTSTSYISKQLEISVPLTSVP-----ELVP 180	
Qy	299 IQLDHHERCDCI 310	
Db	181 VKVANHGGCKCL 192	
RESULT 7		
VBGD_MOUSE		
ID_VBGD_MOUSE	STANDARD;	PRT; 358 AA.
AC_P97946;		
DT 15-JUN-2002 (Rel. 41, Created)		
DT 15-JUN-2002 (Rel. 41, Last sequence update)		
DT 15-JUN-2002 (Rel. 41, Last annotation update)		
DE Vascular endothelial growth factor D precursor (VEGF-D) (c-fos induced growth factor) (FIGF).		
FIGF OR VEGFD.		
GN		
OS Mus musculus (Mouse).		
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
NCBI_Taxid=10090;		
NP		
RP		
RC		
EX MEDLINE=97030234; PubMed=8876195;		
EA Orlandini M., Marconcin L., Ferruzzi R., Oliviero S.;		





CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U00065; AAK68286.1; -  
DR EMBL; U00065; AAL27237.1; -  
DR HSSP; Q63450; 1A06.  
DR NormDep; D1044.3a; CE27894.  
DR WormDep; D1044.3b; CE29743.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR002290; Ser\_thr\_pkinase.  
DR InterPro; IPR002899; WRI/EB.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF01683; EB; 12.  
DR ProDom; PD000001; Euk\_pkinase; 1.  
DR SMART; SM00220; S\_TKc; 1.  
DR SMART; SM00289; WRI; 12.  
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
DR PROSITE; PS00108; PROTEIN KINASE ST; 1.  
DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.  
KW Hypothetical protein; transferase; Serine/threonine-protein kinase;  
KM ATP-binding; Alternative splicing.  
FT DOMAIN 431 703 PROTEIN KINASE.  
FT NP\_BIND 437 445 ATP (BY SIMILARITY).  
FT BINDING 461 461 ATP (BY SIMILARITY).  
FT ACT\_SITE 569 569 BY SIMILARITY.  
FT VARSP\_LIC 1 904 MISSING (IN ISOFORM B).  
FT VARSP\_LIC 905 1012 SPECIOMN/GSDRPVAVVQSHNGYQKQKTKITTEBOIS  
ACICLLAENPAAGHVPRTAAVILRDLVLQYVNMV  
PTITTVVIRQIHVSLAIVTETE -> MEETCSGPK  
SNISFIWHLKLEVPPIIMCLPELLQIVISVAGCP  
GLTFPLFSNENOPCTCPDAGCSYSSGSSRGRTICQVA  
STVNNYICCTSTNO (IN ISOFORM B).  
FT SEQUENCE 1576 AA; 174677 MW; 3A11E573E813498 CRC64;

Query March 5.9%; Score 101; DB 1; Length 1576;  
Best Local Similarity 19.6%; Pred. No. 2.8;  
Matches 43; Conservative 48; Mismatches 76; Indels 52; Gaps 9;

QY 11 KAMATKAKIDKIDKILKLSFCTAK-ETTIRVNRQUTEKEKIFATYSPDK-----GLIS 62  
DB 334 KNTVTDEIDSDMDLQCSIFRVLRLMTFFERSRLD--KVLPTREIPEKFPVGNVSVL 391  
QY 63 RIVNELKQIYKKKKKNPI---KKV--VKDMNRHRSKES-----IYAAKKH 102  
DB 392 TDYQIAKMDNLIENIEIKMDWETVERQAVGEVEYLLDQLGAGAFGCYIVYRKK 451  
QY 103 MKKTSSS---LAIREMQIKTTMRVHLTPVMAIIKKSGNNRDMDEA-GNHHSQOITRT 157  
DB 452 AQSISENPALKLAKK-----IFMTNLNRESDEKSGDMISVKIKIQ 494  
QY 158 KNGTPHVLTHRMILQOSHMTVLSDISELMAKTRIVNL 196  
DB 495 QLRHPNIVRYRRIEVENHRLIYVMDLQGCSLRDLIITM 533

## RESULT 9

VEGC\_HUMAN STANDARD; PRT; 419 AA.  
AC P49767;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Vascular endothelial growth factor C precursor (VEGF-C) (Vascular  
DE endothelial growth factor related protein) (VRP) (Flt4 ligand) (Flt4-  
DE L).  
GN VEGFC.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OC NCB1 TaxID=9606;

RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 103-120.  
RX MEDLINE=96178224; Pubmed=8617204;  
RA Joukov V., Pajusola K., Kaipainen A., Chliov D., Lahtinen I., Kuk E.,  
RA Saksela O., Kalkkinen N., Aittalo K.,  
RT "A novel vascular endothelial growth factor, VEGF-C, is a ligand for  
RT the Flt4 (VEGFR-3) and KDR (VEGFR-2) receptor tyrosine kinases.";  
RL EMO J. 15:290-298(1996).  
RN [2]  
RP ERRATUM.  
RX MEDLINE=96203094; Pubmed=8612600;  
RA Joukov V., Pajusola K., Kaipainen A., Chliov D., Lahtinen I., Kuk E.,  
RA Saksela O., Kalkkinen N., Aittalo K.,  
RL EMO J. 15:1751-1751(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Glial tumor;  
RX MEDLINE=96312526; Pubmed=8700872;  
RA Lee J., Gray A., Yuan J., Luoh S.-M., Avraham H., Wood W.I.,  
RT "Vascular endothelial growth factor-related protein: a ligand and  
RT specific activator of the tyrosine kinase receptor Flt4.";  
RL Proc. Natl. Acad. Sci. U.S.A. 93:1988-1992(1996).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97388482; Pubmed=9247316;  
RA Fltz L.J., Morris J.C., Towler P., Long A., Burgess P., Greco R.,  
RA Wang J., Gassaway R., Nickbarg E., Kovacic S., Charletta A.,  
RA Giamotti J., Flimerty H., Zollner R., Beter D.R., Leak L.V.,  
RA Turner K.J., Wood C.R.,  
RT "Characterization of murine Flt4 ligand/VEGF-C,"  
RL Oncogene 15:613-618(1997).  
RN [5]  
RP SEQUENCE OF 32-41; 112-121 AND 228-233, AND MUTAGENESIS OF ARG-227.  
RX MEDLINE=97377029; Pubmed=9233800;  
RA Joukov V., Sora T., Kumar V., Jeltsch M., Claesson-Welsh L., Cao Y.,  
RA Saksela O., Kalkkinen N., Aittalo K.,  
RT "Proteolytic processing regulates receptor specificity and activity of  
RT VEGF-C.";  
RL EMO J. 16:3898-3911(1997).  
CC -1- FUNCTION: Growth factor active in angiogenesis, and endothelial  
CC cell growth, stimulating their proliferation and migration and  
CC also has effects on the permeability of blood vessels. May  
CC function in angiogenesis of the venous and lymphatic vascular  
CC systems during embryogenesis, and also in the maintenance of  
CC differentiated lymphatic endothelium in adults. Binds and  
CC activates VEGFR-2 (Flk1) and VEGFR-3 (Flt4) receptors.  
CC -1- SUBUNIT: Homodimer; non-covalent and antiparallel.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- TISSUE SPECIFICITY: Spleen, lymph node, thymus, appendix, bone  
CC marrow, heart, placenta, ovary, skeletal muscle, prostate, testis,  
CC colon and small intestine and fetal liver, lung and kidney, but  
CC not in peripheral blood lymphocyte.  
CC -1- PFM: Undergoes a complex proteolytic maturation which generates a  
CC variety of processed secreted forms with increased activity toward  
CC VEGFR-3, but only the fully processed form could activate VEGFR-2.  
CC VEGF-C first form an antiparallel homodimer linked by disulfide  
CC bonds. Before secretion, a cleavage occurs between arg-227 and  
CC ser-228 producing an heterotrimer. The next extracellular step  
CC of the processing removes the N-terminal propeptide. Finally the  
CC mature VEGF-C is composed mostly of two VEGF homology domains  
CC (VHds) bound by non-covalent interactions.  
CC -1- SIMILARITY: BELONGS TO THE PDGF/VEGF FAMILY OF GROWTH FACTORS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

DR EMBL; U58111; AAB02909.1; -
DR HSSP; P15692; 1VPP.
DR Genew; HGNC:12682; VEGFC.
DR MIM; 601528; -.
DR InterPro; IPR004153; CXKC repeat.
DR InterPro; IPR002400; GF_cysknot.
DR InterPro; IPR000072; PD_growth_factor.
DR Pfam; PF00341; PDGF_1.
DR Pfam; PF03128; CXKC; 5.
DR PRINTS; PR00438; GFCYSKNOT.
DR ProDom; PD001629; PD_growth_factor; 1.
DR SMART; SM00141; PDGF_1.
DR PROSITE; PS00249; PDGF_1; 1.
DR PROSITE; PS50278; PDGF_2; 1.
DR Mitogen; Growth factor; Glycoprotein; Signal; Repeat;
  Cleavage on pair of basic residues; Multigene family.
KM CLEAVAGE
KW SIGNAL
FT SIGNAL 1 31
FT PROPEP 32 111
FT CHAIN 112 227
FT PROPEP 228 419
FT DOMAIN 280 362
FT FT
FT REPEAT 280 295
FT REPEAT 304 319
FT REPEAT 328 343
FT REPEAT 347 362
FT DISULFID 131 173
FT DISULFID 162 209
FT DISULFID 166 211
FT DISULFID 156 166
FT DISULFID 165 165
FT CARBOHYD 175 175
FT CARBOHYD 205 205
FT CARBOHYD 240 240
FT MUTAGEN 227
SQ SEQUENCE 419 AA; 4683 MW; 9F598749DB3E014F CRC64;

Query Match
Best Local Similarity 19.8%; Score 96.5; DB 1; Length 419;
Matches 41; Conservative 25; Mismatches 64; Indels 77; Gaps 9;

184 SELMHKTRIVNLMCMYLLTVLDRLNDARKYSCPTPNYSVINREELKANVVF-DR 242
97 ANLNSTRETEIKFAAHNYTEI-LKSIDENMKTKQMPREVCIDVKEGVATNTFFKRP 155
243 CLLVORCGNGC-----CGTVNWSCTGNSGK 269
156 CYSVVRGCGCNSBGLQCMNTSTSYLKTLEITVPLSQPKVVTISFANHTSCRCMSKL 215
270 TV-KKHYEVLQ-----PEGHIR-----RGRAKTMA 295
216 DYYRQVHNSIRSLPAPLPOCAANKTCPTNYMNMNHCICLAQEDFMESDAGDSTDG 275
296 LVDI-----QLDHERCDCICGS--RP 315
276 FHDICGPNKELD-EITCQCVCAGLRP 301

```

```

RN PARTIAL SEQUENCE FROM N.A. (ISOFORM VEGF-B186).
RP STRAIN=Sprague-Dawley; TISSUE=placenta;
RA Mandioca S.J., Pepper M.S.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP PARTIAL SEQUENCE FROM N.A. (ISOFORM VEGF-B167).
RC TISSUE=Heart;
RA Weil J., Eschenhagen T., Miltman C., Scholz H.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Growth factor for endothelial cells. VEGF-B167 binds
CC heparin and neuropilin-1 whereas the binding to neuropilin-1 of
CC VEGF-B186 is regulated by proteolysis (By similarity).
CC -1- SUBUNIT: Homodimer; disulfide-linked. Can also form heterodimer
CC with vegf (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted but remains associated to cells or
CC to the extracellular matrix unless released by heparin (By
CC similarity).
CC -1- ALTERNATIVE PRODUCTS: At least 2 isoforms; VEGF-B186 (shown here)
CC and VEGF-B167, are produced by alternative splicing.
CC -1- SIMILARITY: BELONGS TO THE PDGF/VEGF FAMILY OF GROWTH FACTORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF032925; AAB86884.1; -.
DR EMBL; AF022952; AAB95447.1; -.
DR HSSP; P15692; 1VPP.
DR InterPro; IPR00072; PD_growth_factor.
DR Pfam; PF00341; PDGF; 2.
DR ProDom; PD001629; PD_growth_factor; 1.
DR SMART; SM00141; PDGF_2.
DR PROSITE; PS00249; PDGF_1; 1.
DR PROSITE; PS50278; PDGF_2; 1.
DR Mitogen; Growth factor; Heparin-binding; Alternative splicing;
KW Multigene family.
KM NON TER 1 1
FT FT 47 91
FT DISULFID 51 93
FT DISULFID 41 41
FT DISULFID 50 50
FT VARSPLIC 106 >135
FT FT
FT CONFLICT 29 29 L -> F (IN REF. 2).
FT CONFLICT 37 37 L -> F (IN REF. 2).
FT CONFLICT 98 98 R -> K (IN REF. 2).
FT NON TER 135
SQ SEQUENCE 135 AA; 15001 MW; A915863D8586F82D CRC64;

Query Match
Best Local Similarity 28.8%; Score 95.5; DB 1; Length 135;
Matches 30; Conservative 15; Mismatches 38; Indels 21; Gaps 5;

216 RYSCPTPNYSVINREELKANV--FFPRCLVQRGCGCGTVNWSCTGNSGKTVK 273
13 RATQCPREVVLPSHML-MGNVVKQLVPSCTVYVQCG--GCGPDDGIECVIRIGHQVARM 68
274 VHEVLQPEGHIRKGRRAKTMALVDIQLDHERCDCICGSRRPR 317
69 QILMTQY-----PSSQLGEMSLBHSQCCEC---RPKR 97

```

```

RESULT 10
VEGB RAT
AC 035485; 054881; STANDARD; PRT; 135 AA.
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Vascular endothelial growth factor B (VEGF-B) (VEGF related factor)
DE (VRF) (Fragment).
GN VEGFB OR VRF.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;

```

```

RESULT 11
VEGB MOUSE
AC P49766; 064290; STANDARD; PRT; 207 AA.
DT 01-OCT-1996 (Rel. 34, Created)

```





Db 3257 LEASIEVSKRKYSLIRVNEAIKTEMVNOANLDRSISLVASLPEK-----3303

Qy 256 GTVNRSTCNSGKTVK-----YHEVLQEPGHIKRGAKTMALV 297

Db 3304 --ERMINTTKOPSKTSOELIGNCIISIIYETYPFGHLNERADMLVII 3349

RESULT 14

VEGC ORFN2 STANDARD; PRT; 133 AA.

AC P52584;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Vascular endothelial growth factor homolog precursor.

GN A2R.

OS Orf virus (strain NZ2) (OV NZ-2).

OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;

OC Parapoxvirus.

OX NCBI\_TaxID=10259;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=94076465; PubMed=8254780;

RA Lytle D.J., Frazer K.M., Fleming S.B., Mercer A.A., Robinson A.J.;

RT "Homologs of vascular endothelial growth factor are encoded by the

RT poxvirus orf virus";

RL J. Virol. 68:84-92(1994).

CC -1- FUNCTION: INDUCES ENDOTHELIAL PROLIFERATION.

CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE PDGF/VEGF FAMILY OF GROWTH FACTORS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; S67520; AAB29220.2; -.

DR HSSP; P15692; IVP.

DR InterPro; IPR000072; PD\_growth\_factor.

DR Pfam; PF00341; PDGF\_1.

DR ProDom; PD001629; PD\_growth\_factor; 1.

DR SMART; SM00249; PDGF\_1.

DR PROSITE; PS00249; PDGF\_1; 1.

DR PROSITE; PS50278; PDGF\_2; 1.

KM Mitogen; Growth factor; Glycoprotein; signal.

FT SIGNAL 1 20 POTENTIAL.

FT CHAIN 21 133 VASCULAR ENDOTHELIAL GROWTH FACTOR

FT HOMOLOG.

FT DISULFID 36 78 BY SIMILARITY.

FT DISULFID 67 112 BY SIMILARITY.

FT DISULFID 71 114 BY SIMILARITY.

FT DISULFID 61 61 INTERCHAIN (BY SIMILARITY).

FT DISULFID 70 70 INTERCHAIN (BY SIMILARITY).

FT CARBOHYD 85 85 N-LINKED (GLCNAC. . .) (POTENTIAL).

SO SEQUENCE 133 AA; 14715 MW; 917C0F6883030C39 CRC64;

Query Match 5.4%; Score 92; DB 1; Length 133;

Best Local Similarity 25.0%; Pred. No. 0.81;

Matches 34; Conservative 21; Mismatches 57; Indels 24; Gaps 5;

Qy 193 IYNLMCYTLTVLDR--LNDADKRYSCTPRNSVNIIR-ELKLANVFFPRCLLVORC 249

Db 8 LVAACLHLYLLNAOSNTKGMSEVLKSGECKRPVIVPSETHPELTSGRFNPCTVLMRC 67

Qy 250 GAGCGGCVNMRSCNSGKTVKHYEVLQEPGHIKRGAKTMALVDIQLDHHRCDC 309

Db 68 GAGC---NDSLECVTEEVNVMEL-----GASGSGSNGMGRSLFVHHKDCD 114

Qy 310 -----ICSSRPFR 317

Db 115 RPRFTTTPPTTRPR 130

RESULT 15

VEGC MOUSE STANDARD; PRT; 415 AA.

AC P97953;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Vascular endothelial growth factor C precursor (VEGF-C) (vascular

DE endothelial growth factor related protein) (VRP) (Flt4 ligand) (Flt4-

DE L).

GN VEGFC.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BALB/C;

RX MEDLINE=97164697; PubMed=9012504;

RA Kuk E., Lybounessaki A., Taira S., Kaipainen A., Jeltsch M.,

RA Joukov V., Altalo K.;

RT "VEGF-C receptor binding and pattern of expression with VEGFR-3

RT suggests a role in lymphatic vascular development.";

RL Development 122:3829-3837(1996).

CC [2]

CC SEQUENCE FROM N.A., AND SEQUENCE OF 108-126.

CC STRAIN=BALB/C;

CC MEDLINE=97388482; PubMed=9247316;

CC Fltz L.J., Morris J.C., Towler P., Long A., Burgess P., Greco R.,

CC Wang J., Gassaway R., Nicholas S., Charette A.,

CC Giannotti J., Finmercy H., Zollner R., Beler D.R., Leak L.V.,

CC Turner K.J., Wood C.R.;

CC "Characterization of murine Flt4 ligand/VEGF-C";

CC Oncogene 15:613-618(1997).

CC -1- FUNCTION: Growth factor active in angiogenesis, and endothelial

CC cell growth, stimulating their proliferation and migration and

CC also has effects on the permeability of blood vessels. May

CC function in angiogenesis of the venous and lymphatic vascular

CC systems during embryogenesis, and also in the maintenance of

CC differentiated lymphatic endothelium in adults. Binds and

CC activates VEGFR-2 (Flk1) and VEGFR-3 (Flt4) receptors.

CC -1- SUBUNIT: Homodimer; non-covalent and antiparallel.

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- TISSUE SPECIFICITY: Expression detected in mesenchymal cells of

CC postimplantation embryos, particularly in the regions where the

CC lymphatic vessels undergo sprouting from embryonic veins, such as

CC the perimetaneuric, axillary and jugular regions, and in the

CC developing mesenterium. Expressed in adult heart, brain, spleen,

CC lung, liver, skeletal muscle and kidney.

CC -1- PTM: Undergoes a complex proteolytic maturation which generates a

CC variety of processed secreted forms with increased activity toward

CC VEGFR-3, but only the fully processed form could activate VEGFR-2.

CC VEGF-C first form an antiparallel homodimer linked by disulfide

CC bonds. Before secretion, a cleavage occurs between arg-227 and

CC ser-228 producing an heterotrimer. The next extracellular step

CC of the processing removes the N-terminal propeptide. Finally the

CC mature VEGF-C is composed mostly of two VEGF homology domains

CC (VHDS) bound by non-covalent interactions (By similarity).

CC -1- SIMILARITY: BELONGS TO THE PDGF/VEGF FAMILY OF GROWTH FACTORS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; U73620; AAC52984.1; -.

Query Match	Best Local Similarity	Score 92;	DB 1;	Length 415;
Matches	Conservative	27;	Mismatches 18;	Indels 42;
			Gaps 24;	Gaps 4
DR EMBL; U58112; AAB6707.1; -				
DR HSSP; P15692; 1VPD				
DR MGD; MGI:109124; Vegfc.				
DR InterPro; IPR004153; CXXC_repeat.				
DR InterPro; IPR002400; GF_cysknt.				
DR InterPro; IPR000072; PD_growth_factor.				
DR Pfam; PF00341; PDGF_1.				
DR Pfam; PF03128; CXXC; 5.				
DR PRINTS; PR00438; GRCYSKNOT.				
DR Pfdom; PF001629; PD_growth_factor; 1.				
DR SMART; SM00141; PDGF_1.				
DR PROSITE; PS00249; PDGF_1; 1.				
DR PROSITE; PS02078; PDGF_2; 1.				
KW Mitogen; Growth factor; Glycoprotein; signal; Repeat;				
KW Cleavage on pair of basic residues; Multigene family.				
KM				
FT SIGNAL	1	31		
FT PROPEP	32	107		
FT CHAIN	108	223		
FT PROPEP	224	415		
FT DOMAIN	276	358		
FT				
FT REPEAT	276	291		
FT REPEAT	300	315		
FT REPEAT	324	339		
FT REPEAT	343	358		
FT DISULFID	127	169		
FT DISULFID	158	205		
FT DISULFID	162	207		
FT DISULFID	152	152		
FT DISULFID	161	171		
FT CARBOHYD	171	171		
FT CARBOHYD	201	201		
FT CARBOHYD	236	236		
SO SEQUENCE	415 AA;	46471 MM;	D93JD3CECC659D6 CRC64;	
Query Match	5.4%;	Score 92;	DB 1;	Length 415;
Best Local Similarity	24.3%;	Pred. No. 3.1;		
Matches	Conservative	18;	Mismatches 42;	Indels 24;
			Gaps 24;	Gaps 4
OY 207 LDRLLDNDKARYSCTPRNTSVNIREBLKLANVVF-PRCLLVORCGANGCGTVMNRSCTC	265			
Db 115 LKSIDNEWRKTCQCMREVCIDGKEFGATYNTFFPKPCSVYRCGGCC-----NSEGLQC	169			
OY 266 NSGKTVKKYHEVLQEPGHIRKRGAKMAL-----VDIQLDHERDCDI	310			
Db 170 MATST-----GLSKTLFETITVPLSQGKRPVITSPANTSCRM	208			

Search completed: June 5, 2003, 12:42:21  
Job time : 25 secs

